Education Fights Poverty? A Meta-Analysis of Psychosocial or Educational Preventive Interventions for Children and Adolescents.

Dissertation

zur Erlangung des akademischen Grades

doctor philosophiæ (Dr. phil.)

vorgelegt dem Rat der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena

von Dipl.-Psych. Sabrina Maichrowitz
geboren am 21. Juli 1985

Gutachter

1. Prof. Dr. Andreas Beelmann, Friedrich-Schiller-Universität Jena

2. Prof. Dr. Roland Merten, Friedrich-Schiller-Universität Jena

Tag der mündlichen Prüfung:

09. Mai 2017

Acknowledgements

Writing a dissertation is hardly possible without the support of a lovely family, understanding friends, and supportive colleagues. To all of them I want to express my deepest gratitude.

In particular, I would like to thank my supervisor Prof. Beelmann who supported me with his experienced and critical input whenever needed. He greatly shaped and improved my scientific thinking during the last years.

Furthermore, I would like to thank my research colleagues. Angelika Schulz meticulously checked if all depicted numbers in my dissertation were correct. Louisa Arnold often injected healthy pragmatism into my perspective and Sebastian Schulz was a methodological encyclopaedia. All three of them also proofread parts of my dissertation and provided me with valuable critique. Jutta Proch supported me with her expertise regarding the programme *R*. Without her help, some nice figures would not have made it into the thesis.

Moreover, my gratitude goes to my life partner Pelle Bernhold who endured me during the last year. He built me up whenever I lost confidence and gave me valuable input.

The Friedrich Schiller University of Jena and in particular the Department of Psychology provided me with an optimal working environment and prepared me well on the path of becoming a PhD. Thus, I would like to thank all scientists who made the "Willy" the place it is!

Summary

Since more than 70 years, poverty is scientifically investigated as risk factor for the cognitive and educational development of children and adolescents. Hundreds of programmes have been implemented which aim at preventing and compensating these unequal developmental precondition. The purpose of this dissertation was to synthesize the high-quality evidence worldwide into one comprehensive meta-analysis.

An extensive, multimodal literature search for published and unpublished articles was conducted by scanning electronic databases and thematically related reviews, as well as by contacting authors of primary studies. Studies were included if the authors investigated the effectiveness of a psychosocial or educational programme designed to prevent the negative educational consequences of poverty. Additionally, the sample had to consist of children and adolescents up to the age of 18 years who predominantly (at least 70% of the sample) were of low socio-economic status (low-SES). Furthermore, studies were only eligible if a minimum of 50 participants were randomly assigned to the intervention and the untreated control group.

A total of 109 relevant studies published between 1967 and 2013 could be retrieved. Most of them originated in the United States. The publications provided data regarding 132 contrasts between intervention and control groups (i.e., comparisons). The sample included more than 80 000 children and adolescents predominantly between 2 and 12 years of age with mainly African-American or Hispanic ethnic background. Overall, an unbiased and robust, small but significant mean effect on children's educational development could be found (d = 0.31). Regarding short-term effects (i.e., assessment up to three months after the intervention ended), child and parent trainings were comparably effective (d = 0.35 and 0.27, respectively). Stable long-term effects (more than 12 months post intervention), however, could only be observed if approaching children directly. Combining child and parent trainings can be regarded as promising approach which needs further investigation. Teacher trainings had no significant effect on children's educational development. Regarding the moderating influences of child training effectiveness, structured programmes improved educational development 3.7 times more than unstructured ones. Implementation prob-

lems decreased the effect and so did supervising and observing the trainer during intervention implementation as opposed to well-implemented programmes, as well as unsupervised, unobserved, and untrained ones, respectively. Parent trainings, on the other hand, were 2.5 times more effective if implemented in socially deprived areas than in non-deprived neighbourhoods. Furthermore, if the executors were trained, parent trainings were more than twice as efficient as if they were untrained and 5.7 times more effective if they were manualised as opposed to unmanualised. Moreover, detailed analyses regarding the two child outcome constructs basic and school development were conducted and discussed.

One of the recommendations concerns the outcome structure which is often diluted in intervention research. Future synthesis reviews should derive their outcome structure on the basis of theoretically well thought-through definitions. This dissertation provides one option in how that could be realised. Furthermore, limitations of the thesis are discussed, such as the focus on educational outcomes or the exploratory character of the dissertation.

It can be concluded that the educational development improvement of psychosocial and educational preventive interventions for low-SES children and adolescents is substantial for social policy purpose. However, such programmes are not sufficient for closing the developmental gap between poor and non-poor children and adolescents.

Zusammenfassung

Seit mehr als 70 Jahren wird Armut als Risikofaktor für die kognitive und bildungsbezogene Entwicklung von Kindern und Jugendlichen wissenschaftlich untersucht. Hunderte von Interventionen wurden durchgeführt, um den mit Armut einhergehenden ungleichen Entwicklungsvoraussetzungen vorzubeugen und zu kompensieren. Mit dieser Doktorarbeit soll die qualitativ hochwertige Evidenz, die zu dem Thema vorliegt, in einer umfassenden Meta-Analyse zusammengefasst werden.

Dafür wurde eine umfangreiche, multimodale Literaturrecherche nach publizierten und nicht-publizierten Artikeln durchgeführt, in dem elektronische Datenbanken und thematisch ähnliche Reviews durchsucht und die Autoren der dann eingeschlossenen Primärstudien kontaktiert wurden. Eingeschlossen wurden Studien, wenn die Autoren die Wirksamkeit von psychosozialen und pädagogischen Interventionen untersuchten, die darauf abzielten, negativen bildungsbezogenen Folgen von Armut vorzubeugen. Dabei musste die Stichprobe aus Kindern und Jugendlichen bis zu einem Alter von 18 Jahren zusammengesetzt sein und vorwiegend (d.h. mindestens 70% der Stichprobe) aus finanziell benachteiligten Verhältnissen stammen. Zudem wurden Studien nur dann in die Meta-Analyse integriert, wenn mindestens 50 Teilnehmer einer Interventions- und Kontrollgruppe randomisiert zugewiesen wurden.

Insgesamt wurden 109 relevante Studien eingeschlossen, die zwischen 1967 und 2013 publiziert wurden. Die meisten davon stammen aus den USA. Die Publikationen enthielten 132 Kontraste zwischen einer behandelten und einer nicht-behandelten Gruppe (auch Vergleiche genannt). Daten von mehr als 80 000 Kinder und Jugendlichen mehrheitlich im Alter von 2 bis 12 Jahren und vorwiegend mit afroamerikanischen oder lateinamerikanischen Hintergrund wurden analysiert. Insgesamt konnte ein robuster kleiner, aber signifikanter durchschnittlicher Effekt auf die bildungsbezogene Entwicklung von Kindern und Jugendlichen gefunden werden ($d = 0.31$). Bezogen auf die unmittelbare Effektivität (gemessen innerhalb von drei Monaten nachdem die Intervention abgeschlossen war) waren Trainings mit Kindern verglichen mit Elterntrainings gleichermaßen wirksam ($d = 0.35$ beziehungsweise 0.27). Langfristige Effekte (später als 12 Monate nach der Intervention) konnten

allerdings nur erreicht werden, wenn direkt mit den Kindern gearbeitet wurde. Kinder- und Elterntrainings zu kombinieren scheint vielversprechend zu sein, bedarf aber weiterer Untersuchungen. Lehrertrainings hatten keinen signifikanten Effekt auf die bildungsbezogene Entwicklung der Kinder. Bezogen auf moderierende Einflussvariablen bei Kindertrainings zeigten die Daten, dass strukturierte Programme 3.7 mal so effektiv waren wie unstrukturierte. Probleme bei der Interventionsdurchführung reduzierten die Wirksamkeit genauso wie das Supervidieren und Observieren des Trainers im Vergleich zu gut implementierten Programmen, bei denen keine Supervision oder Observation durchgeführt wurde. Elterntrainings dagegen waren 2.5 mal effektiver, wenn sie in sozial schwachen Vierteln durchgeführt wurden verglichen mit sozial nicht benachteiligten Gegenden. Des Weiteren erzielten Interventionen mit geschulten Trainern mehr als doppelt so hohe Effekte als solche mit ungeschulten und manualisierte Programme waren 5.7 mal effektiver als nicht-manualisierte. Darüber hinaus wurden differenzierte Analysen bezüglich der grundlegenden Entwicklung (wie etwa kognitive Fähigkeiten) und der schulischen Entwicklung (wie etwa Leseleistung) von Kindern und Jugendlichen analysiert und diskutiert.

Eine der genannten Empfehlungen an die Forschung betrifft die Struktur der abhängigen Variablen, die innerhalb der Interventionsforschung oft willkürlich scheint. Die zukünftige Syntheseforschung zu Interventionen sollte die Outcomestruktur anhand theoretisch gut abgeleiteter Definitionen erstellen. Die Doktorarbeit stellt eine Möglichkeit für ein solches Vorgehen vor. Des Weiteren werden Einschränkungen der Dissertation diskutiert, wie etwa ihr Fokus auf bildungsbezogenen Wirkvariablen oder ihr explorativer Ansatz.

Zusammenfassend kann gesagt werden, dass die Verbesserung der bildungsbezogenen Entwicklung durch psychosoziale und pädagogische Interventionen mit präventivem Ansatz für Kinder und Jugendliche aus niedrigem sozioökonomischen Status substantiell für die Sozialpolitik ist. Allerdings sollte dabei nicht vergessen werden, dass solche Programme nicht ausreichen um die Entwicklungskluft zwischen armen und nicht armen Kindern und Jugendlichen zu schließen.

Contents

List of Tables

List of Figures

Abbreviations

95% CI	95% confidence interval
adj	Adjusted
ANCOVA	Analysis of covariance
ANOVA	Univariate analysis of variance
b	Unstandardised regression coefficient
β	Standardised regression coefficient
BD	Basic development
BSV	Between-study variance
χ^2	Statistical test based on the χ^2 distribution
CG	Control group
CO	Child outcome
CSE	Country-specific estimator
CT	Child training
CT/CO	The effect of child trainings on child outcomes
df	Degree of freedom
DSM-IV, DSM-5	Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition, Fifth Edition
Eurostat	European Commission for Statistical Information
F	Statistical test based on the F distribution
FEM	Fixed-effect model
FU	Follow-up
high-SES	High socio-economic status
I^2	Measure of the degree of heterogeneity across studies
ICD-10	International Classification of Diseases 10th Revision
ID	Identification number
IG	Intervention group
κ	Measure for the inter-rater reliability
k	Number of comparisons
low-SES	Low socio-economic status
M	Mean

MANOVA	Multivariate analysis of variance
MEM	Mixed-effects model
MeSH	Medical Subject Heading
N	Total sample size
n_P	Number of publications
obs	Observed
OECD	Organisation for Economic Co-operation and Development
OR	Odds Ratio
PO	Parent outcome
PT	Parent training
PT/CO	The effect of parent trainings on child outcomes
r	Correlation
R^2	Explained variance
REM	Random-effects model
SD	School development
SD	Standard deviation
SE	Standard error
SO	School outcome
TECH	Tool for evaluating research implementation challenges
TNE	Transnational estimator
TT	Teacher training
UOE database	UNESCO, OECD, and Eurostat joint data collection
WSV	Within-study variance
ZPD	Zone of proximal development

1 Introduction and Theoretical Background

1.1 Introduction into Topic

Poverty is a central and tragic topic throughout human history. Since centuries the causes and consequences of poverty are disputed within a variety of disciplines trying to tackle this global challenge from political, economical, philosophical, or religious points of views. However, modern psychological intervention research on poverty began only in 1964 when the United States president Lyndon B. Johnson declared *War on Poverty* by introducing a legislation which was targeted to fight poverty (United States Congress, 1964). Thereafter, the United States' interest in the topic faded, though, as the decline in social welfare expenditures shows (Crouse, Hauan, & Waters Rogers, 2008, p. A-7). Today, poverty is a politically and socially recognised topic which is being addressed within national and international structures. In 1999, for instance, the prime minister of The United Kingdom declared the goal to extinct poverty within a generation (Perry, 2002). Furthermore, the Indian parliament launched the so-called *Right to Food Act* in 2013 which guarantees basic nutrition to most of its poor population (Indian Ministry of Law and Justice, 2013). Since children are particularly vulnerable to the effects of poverty, the United Nations stated in the Convention on the Rights of the Child in 1990 that each child has the right "...to a standard of living adequate for the child's physical, mental, spiritual, moral and social development" (United Nations, 1990, Article. 27(1)). The government in New Zealand announced in 2002 the intention of eradicating child poverty (Perry, 2002). Furthermore, in 2006, most of the member states of the European Union explicated in the *National Reports on Strategies for Social Protection and Social Inclusion* the "...need to develop a strategic, integrated and long-term approach to preventing and alleviating poverty and social exclusion among children" (European Commission, 2008, p. 10). Moreover, the number one *Millennium Developmental Goal* for 2015, which 189 United Nations member states intended to achieve, was to eliminate extreme poverty and hunger (World Health Organization, 2015). This goal was partly attained and has been transitioned into the *United Nations Sustainable Goals* for 2030 to "End poverty in all its forms everywhere" (UN General Assembly, 2015, p. 14).

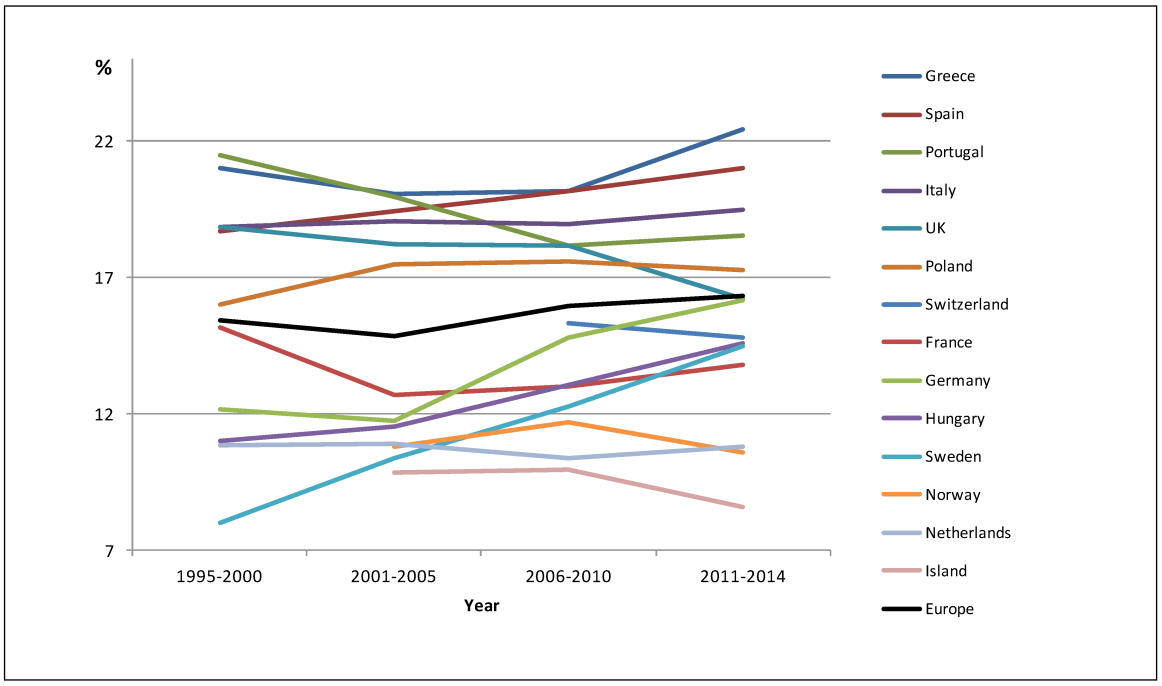
In 2014, 17% of the European population lived below the poverty line (i.e., hav-

ing less than 60% of the median equivalised net income). Hence, poverty affects 85 million Europeans. There are considerable differences across countries, though (see Figure 1). Whereas some have relatively low poverty rates of about 10% (e.g., Island, Norway), others are twice as high (e.g., Greece, Portugal, Spain). Germany ranks in the middle with 17%. However, it belongs to the countries which experienced a 50% augmentation since 2000 (Eurostat, 2016b). In the United States, the situation is even more alarming, with a poverty rate twice as high as in Germany in 2012 (OECD, 2016b)[1]. Across Europe, the poverty rate stayed relatively stable since 1995 which is devastatingly contradictory to the development of wealth: The gross domestic product increased from 1995 to 2013 by 43% (Eurostat, 2016b). The inequality of wealth distribution worldwide is correspondingly shocking, since the richest 10% of the population possessed more than half of the total net value in 2010 (Murtin & d'Ercole, 2015).

[1] The poverty rate is defined as half of the median household income of the total population.

Figure 1

Poverty Rates for Selected Countries in Europe



Notes. Source: Eurostat (code: ilc_li02).

As mentioned above, poverty particularly badly affects the development of children and adolescents, as the numerous evidence shows. Children who grow up in poverty are more likely to live in crowded and unhealthy housing conditions with few stimulating material or opportunities, typically in neighbourhoods with higher crime rates, compared to non-poor children (Bradley, Corwyn, McAdoo, & García Coll, 2001; Dodge, Pettit, & Bates, 1994; Federman et al., 1996; Garbarino, 1999; McCormick & Mason, 1984). Poor parents have limited time for their children as well as fewer financial resources and are, hence, less supportive and patient with their children which manifests itself, for instance, in a more harsh and violent parenting and communication style (Bradley et al., 2001; Dodge et al., 1994; Mabry, 1997; Saxe, Guberman, & Gearhart, 1987). Besides negative effects on the children's socio-emotional development (Bolger, Patterson, Thompson, & Kupersmidt, 1995; Bruce, 1991; Conger, Conger, & Elder, 1997; Dawson, 1991; Dodge et al., 1994; G. Duncan, Brooks-Gunn, & Klebanov, 1994; Felner et al., 1995; Goodman, 1999; Korenman,

Miller, & Sjaastad, 1995; McCoy, Frick, Loney, & Ellis, 1999; Ortega & Corzine, 1990; Sampson & Groves, 1989), poverty already in very early age has severe impairments on children's educational development: Toddlers at 24 months of age are already less attentive, interested, and cooperative (Morgan, Farkas, Hillemeier, & Maczuga, 2009). They quickly fall behind in their cognitive development (G. Duncan et al., 1994; Mpofu & van de Vijver, 2000; J. R. Smith, Brooks-Gunn, & Klebanov, 1997; Stipek & Ryan, 1997) and educational achievement at school (Chase-Lansdale & Gordon, 1996; Conger et al., 1997; Dornbusch, Ritter, & Steinberg, 1991; Felner et al., 1995; Jordan, Huttenlocher, & Levine, 1992; Parcel & Menaghan, 1990; J. R. Smith et al., 1997; Whitehurst & Lonigan, 1998). In addition, problems arise concerning the educational career which expresses itself in more negative attitudes towards school, feelings of being left out, staying down a year, higher drop-out and suspension rates, as well as obtaining lower educational degrees (Brooks-Gunn & Duncan, 1997; Dawson, 1991; Federman et al., 1996; Felner et al., 1995; Prenzel, Sälzer, Klieme, & Köller, 2013; Stipek & Ryan, 1997).

Children are a particularly vulnerable and innocent group. They are randomly born into a poor environment and are victims of their circumstances. If a society agrees to those premisses, all possible measures need to be taken which prevent this devastating development and give children and adolescents from poor families equal chances for achieving prosperous and fulfilled lives.

Since president Johnson's declaration of War on Poverty, the issue became the focus of world-wide public attention. Many programmes were established in the United States which specifically target disadvantaged children. Those interventions are as diverse as the negative developmental consequences of poverty for children and adolescents. They target different parties (e.g., children, parents, teachers, neighbourhoods, a combination of those), are designed based on various developmental theories (e.g., Piaget, Vygotsky, Bronfenbrenner), or aim at improving diverse developmental dimensions (e.g., school achievement, social competence, psychological health), to name only a few dimensions. Many of the programmes which were established as a consequence of Johnson's legislation, were continuously improved over the years and are still implemented today (e.g., Head Start or Job Corps). To the scientifically thoroughly investigated interventions belong the so-called *Early Childhood*

Education Programmes. They usually address children under 4 years of age, work with them directly, but additionally include other parties of the child's direct developmental context, such as the parents or the teachers, and are comparably time intensive. The two most prominent examples are the *Perry Preschool Project* and the *Abecedarian Project*.

Today, we can look back at more than 70 years of intervention research covering hundreds of implemented programmes. In order to shed light on the effectiveness of those programmes, this dissertation will conduct a comprehensive meta-analysis of high-quality studies worldwide which aim at preventing the negative consequences of poverty for children and adolescents.

Before starting, though, the major concepts need to be defined and approaches for their operationalisation have to be explicated (see section 1.2 and 1.4). In this context, the consequences and causal pathways regarding poverty for children and adolescents will be explained (see section 1.3). Then, a closer look at the current state of research regarding this topic will be taken (see section 1.5). Finally, the major research questions will be presented (see section 1.6).

1.2 Defining and Operationalising Poverty

First and foremost, it is important to mention at this point that neither a generally accepted definition of poverty nor a universal way of operationalisation exists. Poverty is a multifaceted concept and its definition and measurement changes due to several factors. One is that the scientific knowledge on the topic grows over time, resulting in adaptations or even new approaches. Furthermore, consumption patterns change, and so do social dynamics and technology. For instance, not owning a mobile phone in the United Kingdom is part of measuring poverty today, whereas 20 years ago it was regarded a luxury good (Sabates, 2008). It further depends on the characteristics of the country which assesses poverty, regarding, for instance, the availability of assessment resources or expertise. Besides those reasonable examples, there are also rather politically and historically motivated explanations for the choice of a definition. Fischer says in his book *Armut in der Geschichte* that, "...concerning poverty we often know more wrong that right..." (Fischer, 1982, p. 8, trans.). He ex-

plains that the discussion "...is not merely motivated by finding the historical truth, but rather driven by the desire for social changes. Moral as well as political impulses are more important, so that history only serves as an illustration" (p. 8).

In addition to keeping in mind scientific, historical, or political aspects when defining and measuring poverty, there is another central aspect that needs to be considered: the availability of data. Since meta-analyses depend on the data reported in primary studies, the definition and operationalisation needs to be oriented toward the available information therein. Keeping all the mentioned aspects in mind, the following sections will define poverty and provide information on how it will be measured in this meta-analysis.

1.2.1 Defining Poverty

Two groups of poverty definitions exist in the literature: absolute and relative poverty. Absolute poverty means possessing less than a fixed minimum standard of living (A. Hagenaars & de Vos, 1988). The minimum standard is absolute in the sense of being invariable concerning different countries. In other words, the same criterion would be applied to a person from Germany and one from Zimbabwe. It is usually understood as a lack of the very basic needs, though, which ensure mere physical survival. One famous example in this category is the definition introduced by the World Bank (1990), which describe poverty as "...inability to attain a minimal standard of living" (p. 26). This definition is implemented by a monetary poverty line of 1$/day at 1993 PPP[2] and was included in the Millennium Development Goals for fighting extreme poverty and hunger worldwide, introduced by the United Nations in 2006. This poverty category should be interpreted as extreme or very severe poverty and does not fulfil the right of a child to a standard of living adequate for the child's development according to Article 27(1) of the Convention on the Rights of the Child (United Nations, 1990). When setting a worldwide absolute poverty threshold, only a minimal part of the society in developed countries would fall below that

[2]PPP stands for purchasing power parity. The amount of 1$ was raised in 2008 to 1.25$/day at 2005 PPP. It is calculated by averaging the national poverty lines of the poorest 15 countries ranked by the per capita consumption (Ravallion, Chen, & Sangraula, 2009).

line. Since the studies included in this meta-analysis will primarily be originated in developed countries and considering the aim to compare interventions of different countries, a relative poverty definition is more suitable. Concerning relative poverty definitions, being poor depends on the social context, in other words, on the level of prosperity in the society or the country, and differs over time. Hence, the central demand is not limited to physical survival, but rather social pressure due to the experienced contrast of ones own living circumstances compared to the ones of others around (UNICEF, 2000). One widespread definition of relative poverty today is the one introduced by the European Union Council of Ministers in 1975, which was then slightly adjusted in 1985 by more emphasising the aspect of social participation:

> *"...'the poor' shall be taken to mean persons, families and groups of persons whose resources (material, cultural and social) are so limited as to exclude them from the minimum acceptable way of life in the Member States in which they live."* (art. 1, part 2)

Shortened but similar is the definition established by the United Nations Economic and Social Council in 1998, stating that poverty "...means lack of basic capacity to participate effectively in society" (p. 1). Those and the many other definitions that exist (e.g., A. B. Atkinson & Marlier, 2010; Citro & Michael, 1995; Encyclopædia Britannica Online, 2015; Galbraith, 1998; Scottish Poverty Information Unit, 2011; Townsend, 1979; United Nations, 2001, 1995) share two general aspects: A lack or deprivation of some kind and the thereby evolving negative consequences. Based on those definitions, relative poverty will be defined as follows:

> *Poverty is understood as a state in which people suffer from relative material deprivation. This can lead to social exclusion and hinders affected individuals and groups to fully participate in society.*

The term *poor*, however, rather describes the concept of absolute poverty. For relative poverty the label *low socio-economic status (low-SES)*, *socio-economically disadvantaged/deprived*, or *low socio-economic background* is more common in the scientific literature and will be subsequently applied.

1.2.2 Operationalising Poverty

Already starting in the 1930s, many attempts to measure SES have been made. Among others, Kahl and Davis (1955) criticised, that the developed approaches were rather practically motivated than scientifically: A high predictive value of the measure seemed to be more important than a theoretical or empirical deduction. One of the consequences was that the new invented indexes proved useful in a given research context which often could not be replicated and lead to "...proliferation and confusion..." (Kahl & Davis, 1955, p. 317). Since then, many studies have been conducted concerning the relationship between the different measures (Blishen, 1958; Bradshaw & Finch, 2003; A. Hagenaars & de Vos, 1988; Kahl & Davis, 1955; Lelkes & Gasior, 2012; Nolan & Whelan, 1996; Perry, 2002; Townsend, 1979). According to the above established definition of poverty, however, low-SES represents a low economic position relative to the society. The economic position can be measured by a variety of concepts. However, some are applied more often and, thus, are widely accepted, compared to others. A. Hagenaars and de Vos (1988), for example, say that low-SES needs to have some connection to income. They justify this thesis by arguing that the concept of low-SES would lack content if the number of people with low-income would be just as much as the number of high-income people. According to Bradley and Corwyn (2002), most low-SES measures can be subsumed to three widespread operationalisations: income, occupation, and education. In this meta-analysis, low-SES will be assessed by the total of those three measures, subsequently referred to as low-income-SES, low-occupation-SES, and low-education-SES.

The most common low-SES threshold in developed countries for measuring material deprivation is set at a percentage of the income (T. Atkinson, Cantillon, Marlier, & Nolan, 2002). Thus, data regarding income is available for a long time period and across countries. Both time and location are important, since the eligibility of the studies in this meta-analysis had no restriction concerning their publication year or the country where the study was conducted (see section 2.1). Hence, it was assumed that in most primary studies (i.e., studies eligible for this meta-analysis) low-SES would be assessed by some form of income threshold. Therefore, low-income-SES was established as the primary criterion and thus a major focus was put on estimating the low-income-SES thresholds for different countries and for several time intervals.

Hence, low-occupation-SES and low-education-SES will serve as additional indices for measuring low-SES.

Low-Income-SES Threshold. Regarding low-income-SES, the most commonly used threshold which also allows to compare different countries, is the one introduced by the European Commission in 2001: 60% of the median equivalised income. As income measure, the disposable household income was chosen because it is equivalised which means it is weighted by the needs of each household member, and it is further expressed per capita (DESTATIS, 2015). Applying an equivalence scale to the income makes it possible to standardise and compare households in terms of their size and age distribution (see Appendix A, Weighting the Thresholds). This threshold is interpreted as an indicator for (risk of) financial poverty and thereby fits the definition in section 1.2.1.

Eurostat collects data for calculating this parameter (subsequently referred to as *threshold table*). Unfortunately, the database only goes back to the year 2000, so that for relevant primary studies published before that date, no thresholds are available. Additionally, only data for the 27 member states of the European Union are on-hand. Hence, in order to judge whether a study was eligible for this meta-analysis, it was necessary to calculate low-income-SES thresholds for desirably all countries starting in the year 1960 (when the first intervention studies with low-SES samples were conducted) until today. First of all, an income database had to be chosen as the basis for the calculations. The most extensive database concerning the number of countries assessed over a long period of time is the one of the Organisation for Economic Co-operation and Development (OECD) regarding disposable income (subsequently referred to as *income table*). Secondly, for the countries represented in both databases, the low-income threshold of the threshold table (containing the desired low-income threshold) was applied on the income table (containing extensive income data). Thereby, a percentage for those countries could be calculated for the year 2002 until 2010. The average over the years 2002 until 2010 within each country then served as its best possible low-income threshold estimator for the time before 2002. For the countries which were only represented in the income table, the estimator was calculated by averaging the above means (from 2002 until 2010 of the

countries available in both databases) across the available countries and also served as best possible low-income threshold estimator for the time before 2002. For those countries which were only represented in the Eurostat-database, the available data was averaged across the years (2002 until 2010), transferred to US Dollars and then directly used as thresholds. At last, the thresholds before 2002 were calculated using the different estimators regarding four time intervals: until 1979, 1980-1989, 1990-1999, 2000-2011. For more detailed information concerning the calculation process, see Appendix A.

The procedure is criticisable in terms of accuracy. What needs to be kept in mind, though, is that the aim was to calculate low-income-SES thresholds that are comparable across countries worldwide and over time. No complete database exists which fulfils all the needed criteria. Thus, the resulting low-income-SES thresholds that are displayed in Appendix B served as advising guidelines rather than strict limits for deciding whether a study was eligible concerning the low-SES criterion.

Low-Occupation-SES Threshold. The most prominent representative of granting occupation a central role for class position was Karl Marx (1818-1883). He saw an individual's role within the production process as the central aspect in determining the individual's position within a class hierarchy. This, in turn, then determines the individual's political and ideological awareness (Rigby, 1987). However, his concept of class hierarchy should rather be interpreted as property relations (i.e., the bourgeoisie or capitalists versus the proletariat or workers) than social relations. Nevertheless, already in Marx' theory a strong association between income and occupation can be found: the bourgeoisie who owns the means of production as upper class with higher income and the proletariat who earns money to survive by producing those means as lower class with less income.

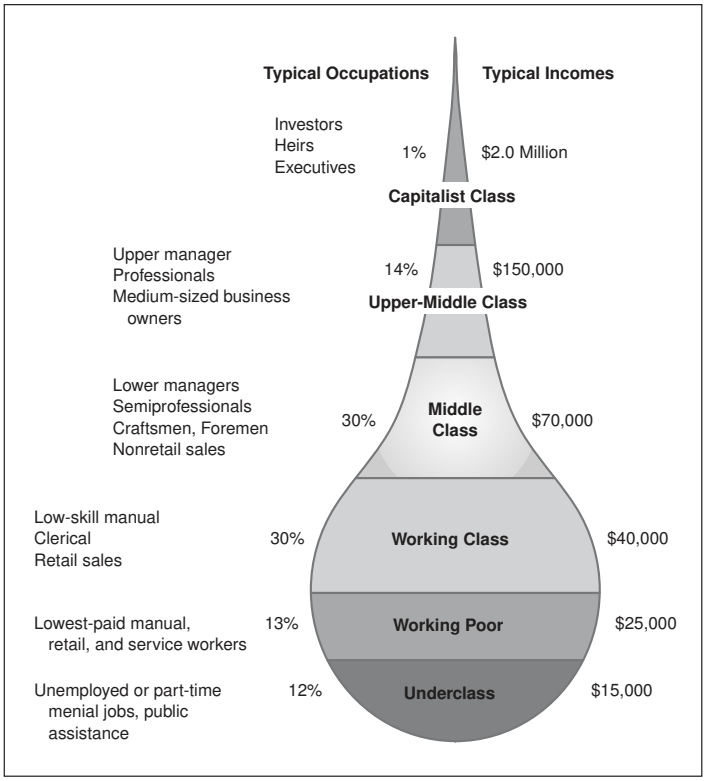
In this thesis, low-occupation-SES will be measured by low occupational status representing a low economic position within a society. That means that pure prestige oriented occupation measures, such as the Siegel scale won't be considered (Miller & Salkind, 2002; Nam & Boyd, 2004). Since low-occupation-SES will serve as an indicator for low-income-SES, it is necessary to directly connect them to each other and, hence, making it possible to apply the low-income-SES threshold on occupation measures.

Most of the established occupation measures for socio-economic status consider income in some form, usually by ranking the occupations based on income, and sometimes also include other information, such as educational status (Blishen, 1958; Blau, Duncan, & Tyree, 1978; O. Duncan, 1961; Nam & Powers, 1983). Lelkes and Gasior (2012) recently analysed the overlap between low-income (i.e., below 60% of the national median equivalised disposable income) and low-occupation (i.e., living in a household where adults worked less than 20% of their total work potential during the past year). They found out that about half of all people living in households with very low work intensity are also income-wise categorised as low-SES. Gilbert (2011) introduced a model which directly compares income and occupation in a common class structure (see Figure 2).

With regard to those publications, the low-occupation-SES threshold based on the low-income-SES threshold was established. Since Gilbert's model was based on data from the United States, the corresponding low-income threshold and equivalence weight for an average American family of two parents and two children in the time interval between 2000 and 2011 is equivalent to a monetary threshold of 33 081 US$ (see Appendix B). Therefore, the low-occupation-SES threshold can be assigned to the lower level of the *working class* in Gilbert-Kahl's model, hence above the *working poor* (see Figure 2). This threshold is supported by the average estimator (see Appendix) which is 42%. Thus, samples rated as low-SES measured by the low-occupation threshold, have jobs which are categorised as unskilled or semi-skilled or, in reference to Lelkes and Gasior (2012), living in households where adults work less then 20% of their total work potential (e.g., part-time jobs, with periods of unemployment).

Figure 2

Gilbert-Kahl Model of Class Structure



Notes. Source: D. L. Gilbert (2001). *The American class structure in an age of growing inequality* (p. 14). Los Angeles, CA: Pine Forge Press.

This threshold is an absolute low-SES threshold and doesn't actually fit the definition of poverty. This exception was made, though, because of economic and empirical reasons. First of all, low-occupation-SES was only included as a secondary criterion. In addition, based on empirical values, it was expected that if intervention studies assessed a sample's low-SES by occupation, it has been likely conducted in the United States. If studies were originated in other countries, the threshold was individually adapted and thereby becoming relative.

Low-Education-SES Threshold. Like establishing occupation as indicator for income capacities, eductaion needs to be related to income, as well. Many statistics exist on education as a general national social indicator. The United Nations, for instance, assess the length of schooling as one indicator of the Human Development Index

since 1990 and publish the results in annual reports (United Nations, 2014). Another example are statistics which are based on data collected by UNESCO, OECD, and Eurostat (also called the UOE database) concerning, among many other measures, educational attainment or the distribution of literacy proficiency (OECD, 2012). In the 1970s, the UOE developed a system for classifying education named the International Standard Classification of Education (ISCED; OECD, 2015). Relevant for this meta-analysis is the distinction of nine educational levels which measure educational attainment ranging from ISCED-0 (early childhood education) to ISCED-8 (doctoral or equivalent level). Those levels correlate high with the median equivalised net income of $\eta = .974$[3] ($N = 30$, $p < .001$), ranging from .509 (Sweden) to .992 (Italy). This supports the approach of using low-education-SES as an indirect measure for low-income-SES. The obvious threshold for low-education-SES appears to be the low educational attainment category (ISCED 0-2) which is also recommended by T. Atkinson et al. (2002): "Where upper secondary education [ISCED 3-4] is required to enter third-level or advanced vocational programmes, then failure to cross the ISCED 2/3 divide is prime facie evidence of exclusion from the possibility of educational advancement along these lines [...][especially], given its role in influencing subsequent life-chances and the risk of experiencing poverty and exclusion" (p. 131). Since the educational levels of the UOE are merely a structural framework, each country can adjust its educational system to it. Thereby, it fits the definition of relative poverty (see section 1.2.1). In conclusion, samples rated as low-SES measured by the low-education threshold, have an educational attainment rated lower than ISCED 3 which, in industrialised countries, corresponds to having competed lower secondary school or being in the educational system for about 10 years.

[3] Included in the calculations were annual data of 27 European countries from 2005 to 2014 on the median equivalised net income separated for the educational levels *low* (ISCED 0-2, preprimary to lower secondary education), *middle* (ISCED 3-4, upper secondary and post-secondary), and *high educational attainment* (ISCED 5-6, first and second stage of tertiary education). The data was retrieved from the Eurostat homepage (SILC [ilc_di08]).

1.3 Consequences and Causes of Low-SES

Since more than 70 years, low-SES is scientifically investigated as risk factor for the cognitive and educational development of children and adolescents (Coleman, 1940). During the last 10 years, the topic moved further into the centre of public attention due to comparative international education studies such as PISA (Prenzel et al., 2013), IGLU (Bos, Tarelli, Bremerich-Vos, & Schwippert, 2012), or Education at a glance (OECD, 2014). Additionally, several reviews have been published during the last 20 years which summarise the evidence about the negative consequences of low-SES for children and adolescents (Aber, Bennett, Conley, & Li, 1997; Arnold & Doctoroff, 2003; Bornstein & Bradley, 2012; Bradley & Corwyn, 2002; Brooks-Gunn & Duncan, 1997; G. Duncan, Yeung, Brooks-Gunn, & Smith, 1998; G. Duncan & Brooks-Gunn, 2000; Evans, 2004; Lee & Burkam, 2002; Leventhal & Brooks-Gunn, 2000; Mayer, 2002; McLoyd, 1998; Yeung, Linver, & Brooks-Gunn, 2002).

The subsequent sections will give an overview of the educational impairments which children and adolescents from low- as opposed to high socio-economic status experience. Furthermore, possible pathways will be explained which might cause this development.

1.3.1 Educational Impairments of Low-SES Children

One major developmental domain with scientifically well-documented evidence regarding the negative influence of low-SES in very young age is cognitive development. Children from low-income families show much lower scores on problem solving, creativity, and memory than children from high-income families (Stipek & Ryan, 1997). Those differences are partly equivalent to one year of developmental delay, as derived from the comparison of disadvantaged kindergartners and advantaged preschoolers. Family income and maternal education are also positively associated with IQ measures: The lower the family income or maternal education, respectively, the lower the measured IQ (G. Duncan et al., 1994; Klebanov, Brooks-Gunn, McCarton, & McCormick, 1998; J. R. Smith et al., 1997). J. R. Smith et al. (1997) reported IQ differences between low- and high-income toddlers at an age already as young as 1 year. Furthermore, low-income children stay behind their high-income pendants

in acquiring basic and intuitive reading skills, called emergent literacy skills, such as verbal fluency (Stipek & Ryan, 1997). Moreover, they do less frequently use taxonomic rather than functional classification strategies which is regarded as a form of cognitive maturation from concrete to abstract thinking (Mpofu & van de Vijver, 2000). Morgan et al. (2009) analysed children's learning behaviour and concluded that, already at the age of 24 months, children whose mother's education is in the lowest quintiles show more learning-related behaviour problems, compared to children of mothers in the highest education quintile. In particular, they are less attentive, more disinterested, and more likely to not cooperate when completing performance tasks with an examiner. Similar results were found for children up to the age of 5 years, as well (S. Campbell & Stauffenberg, 2008). The specified disparities in basic cognitive functions which emerge in very early age, have been shown to predict later school achievement (Badian, 1995; Kurdek & Sinclair, 2001; Stevenson & Newman, 1986; Stipek & Ryan, 1997).

Low-SES further affects the achievement and school performance of children and adolescents (Chase-Lansdale & Gordon, 1996; Conger et al., 1997; J. R. Smith et al., 1997; Whitehurst & Lonigan, 1998). Children from low-income families show much lower scores on reading skills than children from high-income families (Dahl & Lochner, 2005; J. R. Smith et al., 1997; Stipek & Ryan, 1997). The developmental delay is as high as for some of the cognitive domains, precisely, a delay of about one year of schooling (Stipek & Ryan, 1997). Apart from the SES indicator income, both maternal employment characteristics as well as maternal education directly affect the children's verbal facility and reading proficiency (West, Denton, & Germino-Hausken, 2000; Parcel & Menaghan, 1990). According to the statistical report of the National Center for Education Statistics in Washington, D.C., only 38% of the children whose mothers have less than a high school degree, pass the basic reading proficiency level for first-time kindergartners, compared to 86% among children whose mothers hold a bachelor's degree or higher (West et al., 2000). Similar results were found for mathematics performance, with children from low-income families achieving much lower scores on number skills and verbal calculation tasks than children from high-income families (Dahl & Lochner, 2005; Jordan et al., 1992; Stipek & Ryan, 1997). Concerning verbal calculation, the fall back already emerge in very early age: Stipek and Ryan

(1997) discovered that the verbal calculation proficiency of low-income kindergarten children (63 months of age) was equivalent to high-income preschool children's skill (53 months of age). Furthermore, kindergartners whose mothers did not finish high school passed the first three mathematics proficiency levels at lower rates than kindergartners whose mothers have an undergraduate degree (West et al., 2000). Also in adolescent age a negative effect of low-SES on achievement can be observed: Adolescents whose parents are employed in unskilled or semi-skilled occupations achieve lower reading proficiency levels and have lower grade point averaged, compared to children whose parents work in skilled jobs (Felner et al., 1995).

In addition to its negative influence on cognitive development and school achievement, low-SES also affects educational attitudes, skills, and behaviour. Stipek and Ryan (1997), for instance, found that among children with low attainment levels, those from low income families view their achievement more negative and have more negative attitudes towards school than those from high income families. Correspondingly, the sense of belonging at school is lower among adolescents whose parents have an unskilled or semi-skilled occupation as opposed to those whose parents work in a skilled job (Felner et al., 1995). Moreover, Dornbusch et al. (1991) reports that lower parental education is associated with lower grades. Accordingly, children of families who live under the poverty line are about twice as likely to repeat a grade, drop out of high school, and being expelled or suspended from school, compared to children from higher SES (Brooks-Gunn & Duncan, 1997; Dawson, 1991; Federman et al., 1996). Adolescents from low-income families are half as likely as adolescents from high-income families to attend either a 2- or a 4-year-college or to complete the requirements for a bachelor's degree (Federman et al., 1996). Even among adolescents with comparable school performance who fulfil the requirement to study, those with parents who did not complete higher education are less likely to study, compared to those who did (Autorengruppe Bildungsberichterstattung, 2010). In the German school system, it is much less likely that children from working class families attend the Gymnasium, compared to youths whose parents have a high occupational status (Prenzel et al., 2013). According to data collected by the National Science Foundation (2003), adolescents whose father's or mother's attained at least a high school degree are more than twice as likely to obtain a doctoral degree in science and engineering

compared to adolescents whose father's or mother's attained less than a high school degree. In later life, children living below the poverty threshold are almost twice as likely as children living above the poverty threshold to be economically inactive at age 24 (Brooks-Gunn & Duncan, 1997).

1.3.2 Causal Pathways of Low-SES

The above cited evidence clearly gives a picture on how strongly a low-SES family background impairs the educational development of children and adolescents. The logical consequence at this point is to raise questions about the mechanism behind it and explaining the pathways which might lead to those effects.

Home Environment. As mentioned before, the educational impairments can be observed already when children are just about 1 year old. At that young age, the family home is the children's primary environmental context. Thus, several studies have focused on searching for factors within the home environment which explain the influence of SES on child development (Chao & Willms, 2002; Guo & Harris, 2000; Hanson, McLanahan, & Thomson, 1997; Kan & Tsai, 2005; Whitehurst & Lonigan, 1998; Yeung et al., 2002). An obvious explanation is that financial restrictions leave the parents only limited possibilities to provide their children with adequate housing conditions, enriching materials, or supportive parental actions (Brooks-Gunn & Duncan, 1997; G. Duncan & Brooks-Gunn, 2000). Several studies have been conducted that support this thesis as they found a substantial association of family income and home environment (G. Duncan et al., 1998; Garrett, Ng'andu, & Ferron, 1994). Concerning the housing conditions, for instance, families with low-income live under much worse circumstances as opposed to families with a higher income. Their homes are more often crowded (i.e., more than one person lives in each room) and the parents are much more likely not to pay the full amount of the gas, oil, or electricity bill as well as the rent. Correspondingly, they are more often affected by gas, electric, or oil shutdowns and are more likely to be evicted from their home. They are almost only half as likely to own the place which they live in and more than every 4th low-income family expresses the desire to move to another place, compared to only every 10th among higher income families. As a consequence of that, low-income families are more likely to have moved three or more times until the child's 5th birthday, com-

pared to higher income families (Federman et al., 1996). Apart from the inequality regarding basic living circumstances, low- and higher-SES parents also differ in creating and organising their home environment: Homes of children from low-income families contain less enriching materials for supporting the child's cognitive development, compared to homes of high-income families. Musical instruments, for instance, or toys are less probable to be found (Bradley et al., 2001). The same applies to computers and record/tape players (Bradley et al., 2001; Federman et al., 1996; Shields & Behrman, 2000). Furthermore, the number of children's books as well as adults books available at home is lower (Dodge et al., 1994; Federman et al., 1996; McCormick & Mason, 1984). This difference remains throughout infancy, childhood, and adolescence (Bradley et al., 2001). Children of different income levels also spend their leisure time differently. Children from low-income families watch more television and their parents do less often discuss and reflect the seen content with them (Bradley et al., 2001; Federman et al., 1996; McCormick & Mason, 1984). During childhood and adolescence, they are also less often encouraged by their parents to spend their leisure time in organised activities (e.g., sports or creative activities) and their parents are less likely to take their children out for joint activities, such as shopping, a picnic in the park, or to cultural events, such as museum or theatre (Bradley et al., 2001). Hence, children and adolescents living under the official poverty threshold more often participate in unsupervised and unstructured after-school arrangements, such as hanging out or driving around in cars, as opposed to children and adolescents from families at least 200% above the poverty line (Lugaila, 2003; Mahoney & Eccles, 2008). Participating in unstructured leisure activities, in turn, is associated with lower academic and IQ performances (Lord & Mahoney, 2007; Mahoney & Eccles, 2008; Mercy & Steelman, 1982).

Furthermore, low- and high-SES parents interact differently with their children regarding quantity as well as a quality. Parents of low occupational status, for instance, communicate less often with their children than parents of high occupational status: The number of words a 3-year-old child has heard sums up to only 20 million, in contrast to 30 million (Mabry, 1997). Bradley et al. (2001) analysed data of the National Longitudinal Survey of Youth (NLSY), with a focus on the children's home environment. In interview sessions, the interaction between mothers and their children

was observed with the result that, during the visits, mothers with low family income were less likely than higher income mothers to speak to their children, to respond to their children verbally, to encourage their child to contribute to the conversation, and to answer the child's questions/requests with verbal response (Bradley et al., 2001). Moreover, the support in the child's process of learning to read is weaker among low- versus high-SES parents. Mothers who receive a low income, for instance, only read to their children half as often as do mothers with a high income (Bradley et al., 2001). This influence is strong during infancy and early childhood and declines as the child matures into middle childhood (ages six to nine). Furthermore, children of working-class mothers were less often confronted by their mother with complex number activities (e.g., addition with coins or fingers), compared to children of middle-class mothers (Saxe et al., 1987).

When summing up all the evidence cited above, it is not surprising that home environmental differences partly explain the effect of SES on cognitive development of children in preschool age (G. Duncan et al., 1994; Klebanov et al., 1998; Korenman et al., 1995). Furthermore, home environment has been shown to mediate the effect of SES on cognitive and achievement variables when children are in elementary school age (Davis-Kean, 2005; Eamon, 2002; Korenman et al., 1995; Mercy & Steelman, 1982). However, the amount of explained variance by home environment is higher for preschool children's cognitive development than for elementary school children's achievement (G. Duncan & Brooks-Gunn, 2000; Klebanov et al., 1998; J. R. Smith et al., 1997). This leads to the assumption that the influence of home environmental variables decreases with the child's age. This hypothesis was investigated closely by Aikens and Barbarin (2008). They analysed data from the Early Childhood Longitudinal Study (Kindergarten Class of 1998-1999) assessed by the U.S. Department of Education. They concluded that the main factor for explaining initial reading differences between low- and high-SES children, that is beginning to read at kindergarten age, were family characteristics. Those include home literacy environment (e.g., joint book reading, library visits, number of books at home available for the child), parental involvement in the school (e.g., attending parent-teacher conferences, volunteering at school, participating in fundraising), parental role strain and warmth (experiences in the parental role and closeness to child), and attending centre-based

care prior to kindergarten (e.g., nursery school, preschool). However, this effect declines as the children progress in their reading competence and, instead, other factors become more important. Experiences and resources associated with school and neighbourhood conditions (e.g., peers, teacher qualification, garbage on the street, drug selling, violent crime, gang activity) have very few influence on initial reading scores. However, they turn out to be the highest predictor of SES differences during the period of rapid reading growth which is the time span between kindergarten and first grade. Thus, besides home environment, other contexts which influence the children's development nee to be considered. For this purpose, Bronfenbrenner's ecological systems theory can serve as a framework.

Bronfenbrenner's ecological systems theory. According to Bronfenbrenner (Bronfenbrenner, 1979, 1989), a child's development needs to be seen within the context of different interactive environments. The child and the systems change over time and so does the importance of particular systems for the child. As children grow older, out-of-home contexts, such as the school, become more important influences on the child's development. Even though, theoretically, developmental theories, such as the ecological systems theory, have already been included in poverty research since more than 20 years (Aber et al., 1997), there is still very few studies which integrate the complexity of the phenomenon, as, for instance, Aikens and Barbarin (2008) do. This is all the more surprising since inequalities in out-of-home contexts between low- and high-SES children are numerous and well documented.

Educational institution. Besides home environment as developmental context for children, socio-economic differences can also be observed regarding educational institutions. Facilities with a high proportion of low-SES children, for instance, differ from those with lower percentages regarding the setting as well as the teaching staff's attitudes and behaviours. Already among child care centres the classrooms for toddlers less often meet the recommendations concerning adequate ratio and group size in low- versus middle- or high-income centres (Phillips et al., 1994). The caregivers in low-income centres are rated less sensitive and more detachment behaviour can be observed in the interaction with toddlers, compared with middle- or upper-income institutions (Phillips et al., 1994). A similar picture can be found in low-SES schools: They have fewer access to resources and receive less support by parents (Lee

& Burkam, 2002). The schools attended by low-income children are more likely to have security guards and metal detectors, compared to schools primarily attended by high-income children, indicating a less secure school environment (Federman et al., 1996). Furthermore, the support of the teachers is lower in schools with large numbers of students receiving free- or reduced-price lunch as opposed to those with lower percentages of low-income students. Already when hired, new teachers invited to low-income schools more often regard the process as untimely and information-poor, compared to their high-income school colleagues. During familiarisation phase, they receive fewer support by their experienced colleagues, are less often satisfied with their mentor and are more likely to move to another school (Ingersoll, 2002; Johnson & Birkeland, 2003; Johnson, Kardos, Kauffman, Liu, & Donaldson, 2004; Kauffman, Moore Johnson S., Kardos, Liu, & Peske, 2002). Thus, low-income schools have higher proportions of new and less qualified teachers who often teach out of their field, compared to schools that serve low percentages of free- or reduced-price lunch students (Ingersoll, 2002; Lankford, Loeb, & Wyckoff, 2002). In teaching situations, students perform worse if the teacher is distant and disaffected (Alexander, Entwisle, & Thompson, 1987). This association is particularly strong for high-SES teachers working with students whose parents have a low educational status: The teachers rate those students as less mature, having less self-regulatory skills, and hold lower achievement expectations for them which, in turn, negatively influence the students' school performance already after one year of schooling (Alexander et al., 1987; Alexander, Entwisle, & Dauber, 1993; McLoyd, 1998; Rist, 1970). High-SES teachers also perceive the school climate as more negative when low-SES students are in their class (Alexander et al., 1987). Low-income students on the other hand, express their learning state less frequently towards their teachers than their high-income counterparts: They less frequently comment their own performance and competencies, seek for help, look bored, or smile after completing a task (Stipek & Ryan, 1997).

As mentioned already above, low-SES adolescents more often spend time in unsupervised and unstructured after-school arrangements with their peers. Their friendships, however, are less stable (Dodge et al., 1994) since more difficulties within their peer group occur: They are less accepted and respected by other peers, less popular, more often involved in conflicts, and become more frequently a victim of peer

aggression (Bolger et al., 1995; Dawson, 1991; Garbarino, 1999).

Besides school as an out-of-home context, inequalities in neighbourhood conditions influence a child's academic development, as well (Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993; G. Duncan et al., 1994; Leventhal & Brooks-Gunn, 2000). Low-SES families are more likely to live in socially deprived areas where a higher number of violent and household crimes occur which low-SES children are more likely to witness or even experience themselves (Dodge et al., 1994; Federman et al., 1996; Garbarino, 1999). This not only affects the families' perception of their neighbourhood but their behaviour, as well: Less low-income families regard their neighbourhood as safe from crime, compared to higher income families and, correspondingly, the former are more afraid to go out on the street (Federman et al., 1996).

1.3.3 Summarising Statement

The aim of the chapter was to provide the reader with an insight into the consequences and causal pathways of low-SES on selected examples. If one takes all the mentioned pieces of this jigsaw together, children from low-SES families already show delays in their cognitive development and in their (pre)school achievement at very early age which can be partly explained by deprivations in the home environment (e.g., less enriching materials, parental support and communication). Those children are more likely to enter lower quality educational establishment with less motivated and tending caregivers. That is especially crucial in early education because the teachers do not only fill the role as transmitter of knowledge but also as emotional supporter. Hence, instead of buffering the disadvantageous starting conditions, the developmental deficits accumulate even before the children enter school. When they enter school, the initial inequalities are being magnified even more due to lower school quality, as well as prejudiced teacher attitudes and behaviours towards those children. Thus, children growing up in low-SES families are disadvantaged in numerous ways concerning their own development and in their environmental contexts. Those disadvantages emerge long before preschool and are verifiable until university age and beyond. Aikens and Barbarin (2008) summarised the evidence on the topic very suitable as "...ubiquitous, stubbornly persistent, and well documented..." (p. 235). Considering the multiple consequential facets of low-SES and its

temporal dimension, it is not surprising that since more than 60 years numerous and diverse interventions have been conducted which directly or indirectly aim at preventing or compensating the negative developmental consequences of low-SES. The programmes range from one-time instructional meetings (e.g., the teacher explains the importance of bedtime stories to parents) to weekly behavioural trainings (e.g., a nurse shows mothers at home how to best stimulate an infant's development), from concepts of solely working with the child (e.g., tutors help children with their homework) to approaching the parents in order to indirectly affect the child (e.g., teaching parents strategies to enhance their children's school motivation), from programmes which have a very specific content or aim (e.g., reading interventions) to such which aim at improving a wide range of skills (e.g., comprehensive preschool programmes). In order to integrate those diverse programme types into one meta-analysis, it is necessary to explain precisely *what kind* of programmes were of interest. This will be done in the next section.

1.4 Defining the Programmes

The focus of the meta-analysis at hand concerned preventive interventions of psychosocial or educational character which aim at enhancing the educational development of children or adolescents caused by low-SES. Defining the programme will be divided into three parts: programme type, content, and aim.

1.4.1 Programme Type: Preventive Intervention

The term prevention describes a procedure which aims at impeding the development of undesired conditions or events. *Intervention*, on the other hand, will be regarded as a process of active interference with the aim to change an undesired condition and the experiences of a person or a group of people. That includes establishing beneficial environmental conditions, as well (Brockhaus, 2001; Günter, 1995). Another categorisation goes back to Caplan (1964). He distinguished between primary prevention (acting before suffering), secondary prevention (acting during suffering in order to ease the condition), and tertiary prevention (acting after suffering in order to prevent subsequent damage). Independently of the categorisation chosen, the prob-

lem remains that those theoretically quite clear defined concepts turn out to be often inseparable in practice: Preventing reading problems by conducting a preschool reading programme may result in a sample with children who don't yet have problems and others who already lack behind. In order to consider this, the term *preventive intervention* will be applied in this thesis. It is thereby possible to emphasise that conducting a programme in this context is an intervening process of preventive character (Beelmann, 2006).

Furthermore, preventive interventions should be based on theoretically substantiated and empirically validated work and the intervention aims should be explicated before conducting the intervention. They should contain a positive approach in the sense of aiming for specific or general developmental enhancements, instead of merely aiming at avoiding negative or undesired attitudes or behaviours (Brandtstädter, 1982; Beelmann & Raabe, 2007; Coie et al., 1993; Heinrichs, Saßmann, Hahlweg, & Perrez, 2002; Perrez, 1994).

1.4.2 Programme Content: Psychosocial or Educational Character

Psychology primarily addresses internal experiences and external behaviour of individuals while for sociologists the group forms the smallest entity to study. Thus, the focus lies on coexisting and collective behaviour (Brockhaus, 2001; Dorsch, Häcker, Stapf, & Becker-Carus, 2009). Shifting from a psychological to a sociological perspective is accomplished by expanding the scientific subject. However, this does neither mean that sociology contains the psychological approach, nor that less importance is attributed to the latter. In order to meet the holistic aspiration of this meta-analysis, the described shift is essential: Moving away from a solely individual psychological approach for the purpose of regarding the children and adolescents both as individuals and, as Bollnow (1963) states, "...within [...] their interlacement with society..." (p. 71, trans.). This implies the integration of the psychological and sociological perspective into a psychosocial approach.

In contrast to psychology and sociology, pedagogy defines itself not primarily by the subject (be it individual or group, respectively), but rather by the kind of acting and influencing (Dorsch et al., 2009). Hence, it is of educative and instructional character and, in this sense, more than psychology and sociology, an intervening and

normative science by definition. In order to demarcate pedagogy from psychology, Bollnow (1963) writes that psychology is concerned with "...how humans are by nature, but not how they should be shaped by education..." (p. 70, trans.).

In practice, psychological, sociological, and pedagogic elements of an intervention are hardly separable, as it is the case for prevention and intervention. The interdisciplinary combination of the three scientific approaches realised in this meta-analysis broadens the scope and thereby makes it possible to represent the natural circumstances of the children and adolescents more precisely. Hence the preventive intervention need to be of psychosocial or educational character. Concretely, the psychosocial or educational approach will be defined by the methods applied during conduct. Typical methods are, for instance, psycho-education, discussions, guided play, exercises, or behavioural training. Those methods need to be applied in a social interaction between the trainer and the trainee (Perrez & Baumann, 2011). The addressee could be the child or to the child related parties, such as its parents or teachers.

1.4.3 Programme Aim: Educational Development

The meta-analysis at hand focuses on preventive interventions which aim at directly or indirectly improving the educational development of children or adolescents from low-SES. The aims can, thus, range from the child's cognitive development and school achievement to parental or school support. Correspondingly, the assessed outcomes are diverse and a systematisation is necessary. There will be one primary and two secondary outcome categories. The primary outcome category contains all child-related outcomes and the secondary category all outcomes which focus on changing the direct living environment of the child.

1. Primary outcome category: Child outcomes
    - Basic development
    - School development
2. Secondary outcome categories: Parental and school support

The outcome structure of the primary outcome category is based on the theories of the psychologists Jean Piaget and Lev Vygotsky. Piaget views humans as organisms

who search for knowledge (Kesselring, 1999). He introduced the idea of children as lone learners who have an aspiration to knowledge and learn best when making their own experiences driven by their own efforts, irrespective of others. Hence, he believed development to be a spontaneous and natural process. He contrasted it to teaching as a standardised school-related practice which he saw as external constraint to development. Vygotsky (1978/1997) recognised the importance of this natural and spontaneous development but saw assisted learning as evenly important for the child's development. He introduced the zone of proximal development (ZPD) which defines the developmental potential of a child by the distance between the child's ability to independently solve problems and to solve problems under guidance of adults (or with the help of peers). From his point of view, teachers are supposed to present tasks to students which are too difficult to be solved alone and guide them in solving them.

Thus, within the primary outcome category a distinction between basic development and school development is established. Basic development corresponds to Piaget's concept of natural, school-independent development. School development, on the other hand, is based on Vygotsky's teaching-related and support-oriented developmental concept. Also included in this construct was the child's development throughout life in institutional settings, including variables such as job satisfaction or income. However, it was expected that the great majority of the relevant intervention studies focussed on school-related outcomes, hence, the category is labelled school development (and not, for instance, institutional development).

The secondary outcome category consists of the educational support which children and adolescents receive by their parents and school. Parental support refers to the assistance by parents, containing all measures which assess directly (e.g., reading of parent to child) or indirectly (e.g., number of books at home) parental attitude, knowledge, or behaviour concerning the child's educational development. School support includes all measures which concern the school as a learning environment and its teaching staff, such as teacher-student interaction, teachers' proficiency, or the classroom climate).

1.4.4 Summarised Definition

Psychosocial or educational preventive interventions which aim at enhancing the educational development of children or adolescents are defined as

a) an intervening process of preventive character,

b) with contents that are based on theoretically substantiated and/or empirically validated work,

c) using methods which focus on internal human experiences and external behaviour (psychological approach), are related to coexistence and social behaviour (sociological approach), or are of educational nature (educational approach), such as psycho-education, discussions, or exercises, and

d) aim at preventing the child's or adolescent's negative educational development caused by low-SES. That includes direct child promotion (i.e., promoting the child's basic or school development), as well as indirect child promotion via parental or school support.

Since all relevant constructs are now defined and operationalised, it is important to investigate whether similar meta-analyses already exist and what conclusions they made.

1.5 Current State of Research

Since the 1960s much research on programme effectiveness was conducted in order to improve the educational chances of children with low-SES. Consequently, many reviews and meta-analyses did focus on that topic. In Table 2 the research of the last 10 years is summarised and will be discussed in this section. The table gives an overview of central study characteristics and the reported effects. First, general information, including the synthesis method used by the author (i.e., *type*), the country where the programmes were conducted (i.e., *nation*), and the time span in which the analysed studies where published or conducted (i.e., *span*), are listed. Subsequently, the kind of intervention and the included outcomes are described in more detail (i.e., *interventions, outcomes*) as well as the study design (i.e., *design*) which the primary studies

implemented. Important to notice at this point is the differentiation between a control group and a comparison group. The former includes only untreated participant, that is a sample which did not participate in an intervention during the investigation. In contrast to a control group, a comparison group is of broader character, since it involves untreated groups as well as such with an alternative intervention. The two then following variables in Table 2 concern sample characteristics. First, the number of included studies (i.e., *k*) and the proportion of low-SES samples (i.e., *disadvantaged*) are displayed. Then, the samples age (i.e., *age*). The table concludes with an overview of the observed effects (i.e., *effects*).

Table 2

Overview of Meta-Analyses and Reviews in This Field Published After 2005

Citation	Type	Nation	Span	Interventions	Outcomes	Design	*k*; Disadvantaged	Age	Main Effects
Kim & Quinn (2013)	MA	US, CA	1998-2011	Summer reading programmes	LA	CtrGD	35; 21 studies w/ ≥50% low-SES	K-8th grade	Overall: d = 0.10* Low-SES: d = 0.10*
Camilli et al. (2010)	MA	US	1960-2000	Centre-based early education interventions	CD, LA, SED	CompGD	123; Almost all studies w/ low-SES[c]	3-5 y.	CD: d = 0.23**[d] SA: d = 0.14**[d] SED: d = 0.16**[d]
Burger (2010)	R	WW	1990-2010[a]	Centre-based early childhood interventions	CD	CompGD	32; 8 studies w/ low-SES[c]	0-5 y.	Overall positive effect
Manning et al. (2010)	MA	WW, all US	1970-2008	Early developmental prevention programmes	ES, CD, SED, D, SP, CJ, FW	CtrGD	17; All studies w/ economically, educationally, ethnically, or linguistically disadvantaged[c]	0-5 y.	Overall: d = 0.31*** ES: d = 0.52*** CD: d = 0.34*** SED: d = 0.16** D: d = 0.48*** SP: d = 0.37*** CJ: d = 0.24*** FW: d = 0.18**
Darrow (2009)	MA	nr	1990-2007[b]	Curriculum interventions	LA	Pre/post CtrGD	17; All studies w/ ≥50% low-SES	3-5 y.	Overall: d = 0.07

continued on the next page...

Table 2

Overview of Meta-Analyses and Reviews in This Field Published After 2005 (continued)

Citation	Type	Nation	Span	Interventions	Outcomes	Design	*k*; Disadvantaged	Age	Main Effects
Lauer et al. (2006)	MA	US	1985-2003	Out-of-school-time programmes	SA	CompGD & CtrGD	35; All at-risk (e.g., low-SES, racial or ethnic minority, single-parent)[c]	K-12th grade	Reading: d = 0.13* Mathem.: d = 0.17*

Notes. Type = type of publication, MA = meta-analysis, R = review, Nation = country where the programmes were conducted, WW = worldwide, nr = not reported, Span = period of time in which the included studies were published, DV(s) = dependent variable(s), LA = literacy achievement, CD = cognitive development, SED = socio–emotional development, ES = educational success, D = deviance, SP = social participation, CJ = involvement in criminal justice, FW = family well-being, SA = School achievement, CtrGD = Control group design, CompGD = Comparison group design, *k* = number of studies included in the report, w/ = with, K = kindergarten, y = years, .

[a] Already reviewed articles were excluded.

[b] No information reported, so year of last included study is noted.

[c] Threshold not reported.

[d] Unweighted effect size

*$p < .05$, **$p < .01$ ***$p < .001$

The most recent meta-analysis was conducted in 2013 by Kim and Quinn. They investigated the effectiveness of summer reading interventions on literacy achievement of children from kindergarten to eighth grade. The intervention needed to be either home- or classroom-based. They included studies which implemented an experimental or quasi-experimental control group design and restricted the publication date to 1998. Thirty-five articles were integrated. Even though the authors intended to investigate low-income children, only 21 of the 35 studies had a sample of at least more than 50% low-SES children. The remainder had mixed or unknown status. The mean effect size on reading achievement across all studies was small but significant, yet heterogeneous. A similar result was found for the low-SES sub-sample. It is worth mentioning, that only studies conducted in the United States or Canada were included and children younger than kindergarten age and higher than 8th grade were excluded.

Camilli, Vargas, Ryan, and Barnett conducted a meta-analysis in 2010 synthesising centre-based early education interventions which provide educational services directly to children. The interventions needed to target the children's cognitive and/or language development. Social outcomes where included if available. The studies had to have a quasi-experimental or randomised design with a comparison group. Even though Camilli et al. did not explicitly include low-SES children, almost all families came from low income backgrounds. It is unclear, though, at which threshold a sample is considered low-income. One hundred and twenty-three comparative studies could be included. Early education intervention showed no effect on the three outcome categories when compared to an alternative intervention. However, it resulted in small but significant effects on all three outcome domains when compared to an untreated control group. The latest included studies were published in 2000, though, and the country were the programmes were implemented was restricted to the United States. Furthermore, the age of the children was restricted to 3-5 years. Only programmes were considered which provided services for at least 10 hours per week for two months.

Recently, Burger published a review in 2010, summarising centre-based early childhood interventions which focused on the promotion of the children's cognitive development. A comparison group had to be involved and a study sample of at least

300 included. Burger included children from different social backgrounds. Thirty-two studies were included. Of those only eight targeted socio-economically disadvantaged children. Overall, it was concluded that the interventions had "...considerable positive short-term effects and somewhat smaller long-term effects on cognitive development and that in relative terms children from socio-economically disadvantaged families made as much or slightly more progress than their more advantaged peers" (p. 140). It should be critically mentioned that only studies published after 1990 were considered relevant and only if they were not already included in other existent reviews. This was justified by the aim not to duplicate existing reviews. This approach should be regarded critical since meta-analyses never conduct exactly the same research, even if they access the same study pool. The case here is that, first, different coding handbooks were used, thus, other kinds of variables retrieved and analysed. Second, the inclusion criteria differed which puts the alleged duplicates in a different context. Furthermore, excluding studies with less than 300 participants lead to a study pool with only non-randomised quasi-experimental designs. This decision was motivated by the intent to analyse "...more typical (real-world) experiences of children..." (p. 144) and, therefore, achieving a high external validity. However, if the observed effects cannot surely be traced back to the intervention and other factors may account for the assumed effect, it is a threat to internal validity. In this case, there is no reliable result that could be generalised. In other words "...jeopardizing internal validity for the sake of increasing external validity usually entails a minimal gain for a considerable loss" (Cook & Campbell, 1979a, p. 84).

The meta-analysis conducted by Manning, Homel, and Smith (2010) reviewed the effectiveness of programmes which focused on developing or enhancing child, parent-child, or family well-being. The authors included a wide range of outcomes (educational success, cognitive development, socio-emotional development, deviance, social participation, involvement in criminal justice, and family well-being, but restricted the assessment time-point to having measured at least one outcome during adolescence (post or follow-up, respectively). The studies needed to have a randomised or matched control group. No restriction was made concerning the country were the studies were conducted. Yet, all included studies were originated in the United States. The programmes had to be directed at disadvantaged and/or low-SES

populations, meaning that the sample could consist of being disadvantaged because of economic, educational, ethnic, or language reasons. The percentage of how many families in each sample live under those circumstance is not reported. Across all programmes (k = 17), a moderate effect could be found, ranging from small to high effects, depending on the outcome. No separate analyses were conducted differentiating the various disadvantaged groups.

Catherine L. Darrow performed a meta-analysis in 2009. She included curriculum-based preschool intervention programmes that were carried out in school settings with children from low-income families between the ages of three and five. Studies were only considered, if they focused on language and literacy development and employed a pretest/posttest control group design. Over all 17 studies were included and no effect of the interventions on vocabulary outcomes could be found. Yet, the significant heterogeneity among the studies, suggested the existence of (a) moderating variable(s). Darrow found predominant ethnicity (more than 50% white versus other) to be such a moderator: For children of non-white ethnicities (African American, Hispanic or Asian/Pacific Islander) the preschool programmes were more effective, in terms of vocabulary development, than for children of predominantly white ethnicity. Even though Darrow focused on investigating children from low-income backgrounds, she considered it already sufficient that more than 50% of the sample were low-income families. As in the case of Kim and Quinn (2013), the heterogeneous sample could be the source for the high heterogeneity among the studies. Besides that, only studies completed after 1990 were included and the age range for the children included, must be judged rather narrow (3-5 years).

Lauer et al. (2006) conducted a meta-analysis regarding the effects of out-of-school-time programmes (e.g., after-school programmes and summer schools) on the development of reading and/or mathematics achievement. To be included, the authors had to employ a control or a comparison group and the studies had to be published after 1985. Children and adolescents at risk were included whose age ranged from kindergarten age to 12th grade. *At-risk* was defined rather broad, containing children from low-SES as well as from racial or ethnic minority background, single-parent families, or also mothers with low education or limited English proficiency. Included in the analysis were 35 studies, yielding small but statistically significant positive ef-

fects on reading and mathematics achievement. Like Camilli et al. (2010), Lauer et al. also restricted the studies to such implemented in the United States. The latest one considered was published in 2003.

1.5.1 Summary and Objectives of the Thesis

The existence of numerous primary studies, meta-analyses, and reviews regarding this topic, shows its social, political, and scientific importance. However, the present meta-analyses and reviews often solely focus on certain aspect. Hence, the evidence is scattered throughout the different papers: They either focus on rather specified programmes (Kim & Quinn, 2013), restricted the sample to a certain age group (Kim & Quinn, 2013; Camilli et al., 2010; Darrow, 2009), included a rather heterogeneous sample by considering different types of disadvantaged populations (Burger, 2010; Manning et al., 2010; Lauer et al., 2006), or limited the country of origin to the United States (Camilli et al., 2010; Lauer et al., 2006).

The dissertation at hand will comprise the scattered evidence into one meta-analysis focusing on low-SES children and adolescents. It will be a comprehensive analysis of effectiveness studies of psychosocial or educational preventive interventions which aim at improving the educational development of children and adolescents worldwide. The stability of the effects will be looked at, as well as the influence of different training types and outcome categories. It will further be analysed to what extent the effectiveness varies regarding characteristics of the intervention or the sample. Moreover, child outcome effectiveness will be looked at in detail to meet the aim of focussing on the developmental changes of children and adolescent.

1.6 Research Questions

This meta-analysis will address the following research questions:

1. **Publication Bias:** The distortion of the overall mean effect by selective publication processes will be analysed.

2. **Overall Effectiveness:** The overall mean effectiveness will be investigated, as well as its robustness regarding alternative decision in the process of conducting

this meta-analysis (i.e., sensitivity analysis).

3. **Long-term Effectiveness:** The stability of the intervention effect over time and by outcome category will be looked at.

4. **Short-term Effectiveness by Intervention Type:** The short-term effectiveness (including only outcomes which were assessed three months after the intervention ended) will be analysed by different programme types. It will be distinguished between child trainings, parent trainings, teacher trainings, and a combination of child and parent trainings.

   Many meta-analyses focus on a certain type of interventions, such as child trainings (Burger, 2010; Camilli et al., 2010; Darrow, 2009; Lauer et al., 2006) or parent trainings (Astuto & Allen, 2009; Brooks-Gunn & Markman, 2005). Some include diverse programme types, however, they primarily investigated the overall effect and moderating variables but did not separately present the results, for instance, for child versus parent trainings (Kim & Quinn, 2013; Manning et al., 2010). This meta-analysis focusses on the promotion of the educational development of children and adolescents, and by that including a wide range of programmes. Thereby, it will be possible to compare the effectiveness of different training types, that is child versus parent versus teacher versus combined.

5. **Moderator Analyses:** Moderating variables will be investigated by training type and outcome category. The potential moderators will be separated into three categories, namely: methodological moderators, sample moderators, and intervention moderators. Some of them will be analysed in more detail:

   - **Intensity of the Intervention:** Since the influence of an intervention's intensity on its effectiveness is controversially discussed in social science, variables regarding this topic will be investigated in detail. Hereby, it will be distinguished between *length* and *density* (see Appendix H).
   - **Implementation Fidelity of the Intervention:** Assessing variables concerning the implementation of the intervention has only recently become an issue in intervention research (Durlak & DuPre, 2008; Griffin, 2010). With regard to Durlak and DuPre (2008), implementation was defined as

"...what a programme consists of when it is delivered in a particular setting" (p. 329). More precisely, certain aspects of implementation, according to criteria published by Dane and Schneider (1998), will be investigated. Among those is *fidelity of the programme*, which assesses to what extent the programme has been conducted as planned. Assuming that the intervention was based on theoretically substantiated and empirically validated work (see section 1.4), high implementation fidelity should result in higher effect sizes. Fidelity will be investigated by the variables *manualisation* and *structuredness* of the programme, as well as *supervision*, *observation*, and *training* of the executor. Furthermore, *implementation problems* will also be analysed as moderator since they are expected to lead to a limited realisation of the intervention content and, hence, result in lower effectiveness.

6. **Differentiated Outcome Analyses:** Since the educational development of children and adolescents was the centre of interest in this meta-analysis, training effectiveness for basic and school development will be analysed in detail, by contrasting different training types. Furthermore, the influence of selected moderators will be investigated. The selection will depend on the results regarding moderator analyses on outcome category level (see research question 5).

The structure of the research questions will be followed when reporting the results (see section 3) and when discussing the findings (see section 4).

2 Method

When conducting a meta-analysis, five stages need to be passed according to Cooper, Hedges, and Valentine (2009): Formulating, Finding, Coding, Analysing, and Publishing. Formulating describes the original step of getting an overview of the current state of research and, on that basis, phrasing the research question and defining the eligibility criteria for the primary studies. An overview of the current research is given in section 1.5 and the research questions are specified in section 1.6. The eligibility criteria will be described in detail in section 2.1. After this initial step, relevant studies need to be searched for and retrieved (see section 2.2). Subsequently, coding the studies follows: Relevant information for answering the research questions have to be extracted from the primary studies by means of a coding handbook which was developed in advance for this purpose (see section 2.3). In the penultimate phase, standardised effect sizes are calculated in order to compare and analyse the results of the primary studies based on explicitly stated assumptions (see section 2.4 and 2.5). After interpreting the results, they have to be summarised and conclusions to be drawn for communicating them to research and practice (see section 4).

It should be mentioned that this meta-analysis was conducted in the context of a research project named PrAMi which started in 2011, financed by the Federal Ministry of Education and Research (grant number 01JC1105). PrAMi stands for preventing negative consequences of poverty and migration. It started in October 2011 and was lead by Prof. Andreas Beelmann. Within the project, two comprehensive meta-analysis with common eligibility criteria were conducted, except for the study population which either was of low-SES (PrAMi-Poverty) or a sample with migration background (PrAMi-Migration). Whereas the PrAMi-Poverty part was realised by the author of this dissertation, Dipl.-Psych. Sebastian Schulz took care of the PrAMi-Migration part. In July 2013, Louisa Arnold joined the project as a third PhD student.

2.1 Eligibility Criteria

The studies included in the meta-analysis at hand had to meet the following eligibility criteria:

1. **Intervention Criterion:** Preventive psychosocial or educational interventions with an educational objective had to be conducted.
2. **Population Criterion:** Children or adolescents without clinical diagnosis from low-SES backgrounds had to be targeted.
3. **Methodological Criterion:** Random assignment with a minimum of 50 participants to at least one intervention and control group had to be applied.
4. **Outcome Criterion**: At least one educational outcome regarding children or adolescents had to be assessed.
5. **Effect Size Criterion:** Effect size calculation had to be possible.
6. **Duplicates Criterion:** Duplicates were included but summarised under one main study.

The criteria will be described in more detail in the following sections. For inclusion, the studies had to be published or otherwise made accessible until the end of 2013. In order to ensure a high generalisability of the results, no restrictions were set concerning the publication language or the country where the study was conducted.

2.1.1 Intervention Criterion

The study had to investigate the effectiveness of preventive psychosocial or educational interventions designed to prevent the negative educational consequences of low-SES as described in detail in section 1.4.4. Furthermore, studies were only included if the authors explicitly mentioned (e.g., within the theoretical deduction of their research question) that they aimed at improving the educational development of children or adolescents. This could be done by directly approaching the children (e.g., child's cognitive development and school achievement) or indirectly by working with the parents or teacher (e.g., parental or teacher school-related support).

Reasons for Exclusion

- **Aiming to Improve Physical Health:** Programmes were excluded which aimed at improving physical health, such as changing the nutritional behaviour

(e.g., preventing underweight or overweight nutrition behaviour, prolonging the breastfeeding period), improving personal hygiene (e.g., supporting hand washing), preventing diseases or disease-related behaviour (e.g., increasing the immunisation rate, decreasing the AIDS/HIV rate), or decreasing the risk of injuries.

- **Aiming to Improve Psychosocial Skills:** Programmes that primarily aimed at improving the psychosocial skills of children or adolescents, such as decreasing the risk of depression, anxiety, aggressiveness, hyperactivity, delinquency, substance abuse, or pregnancies but also improving social interactions, empathy, communication skills, self-esteem, or self-confidence, were excluded.

- **Aiming to Change Interpersonal Relationship:** If the authors of a programme primarily intended to improve the interpersonal relationship between parents and children (e.g., parenting style, parental sensitivity or warmth, parental abuse), the study was excluded.

- **Marginal Psychosocial or Educational Content:** If a programme only marginally contained psychosocial or educational contents, the study was excluded. This was the case if a desired behaviour was intended to be changed, for example, by regular drug tests or detention. Studies which let children listen to classical music in order to improve their mathematics achievement were also judged as marginally psychosocial or educational content.

- **Availability of Material:** Studies which primarily provided the children or adolescents with material (e.g., distribution of books, computers) were excluded, unless the material contained interactive psychosocial or educational contents, such as interactive e-books or educative and interactive computer programmes.

- **Monetary Programmes:** Programmes which applied primarily financial means, such as the usage of financial incentives or cuttings for supporting a desired behaviour, were not included in the meta-analysis. This includes micro-loans, earning or food supplements, conditional cash transfers, scholarships, subsidised or public housing, social welfare, unemployment benefits, or subsidies for day care.

- **Religious Contents:** Studies were excluded which implemented contents based on religion, such as prayer or confession.
- **Judicial Contents:** Programmes that were designed for the judicial system, such as changing the way defendants are being treated, the imposition of a milder punishment, so-called restorative justice programmes, or rehabilitation programmes, were not included in the meta-analysis.

2.1.2 Population Criterion

Age of Participants. The target population were children or adolescents. They were first and foremost defined by an average age of under 18. This threshold was set because of the transition period to early adulthood which begins in many countries at that age (Valsiner & Connolly, 2003). However, besides the factual age, the developmental age was also taken into account which means that some studies were included, even though the average age of the sample exceeded 18 years. This was, for example, the case for young adults who stayed down a year and, therefore, were still in school and probably lived with their parents. Their developmental tasks were judged rather juvenile and, hence, the study was included. As opposed to that, the developmental tasks of teenage mothers were considered rather adult (even though they often live with their parents). Consequently, a study with teenage mothers was included if the authors considered them as parents (e.g., conducting a maternal skills training), but excluded if they were regarded as teenagers (e.g., programme for preventing further pregnancies).

If the main addressee of the programme were not children or adolescents but rather persons relevant in their life, such as their parents or their teachers. This study was only included if the authors stated that the programme aimed at primarily reaching the children (by working with a third party). When training the parents to read aloud to their children or improving the parental skills, for instance, the study was included. The same does apply for programmes which train teachers to use a certain school curriculum.

Participants With Low-SES. At least 70% of the sample had to be of low-SES in the sense of being materially deprived compared to other populations in the country

they live in (see section 1.2). The study authors needed to aim at preventing negative consequences of low-SES (e.g., by mentioning it in the theoretical deduction of the research question in the study paper) and explicitly describe characteristics concerning the low-SES sample. The percentage of 70% was determined in order to obtain a rather homogeneous sample with a great majority of low-SES participants. If instead of the whole sample, a subgroup sample was eligible, only the subgroup was included, as long as the study authors reported enough information for calculating an effect size.

Several criteria were specified for measuring low-SES. They are ordered in a hierarchical manner: The primary criterion for material deprivation was low-income. If no information concerning the income status of the sample was reported, the secondary criteria were applied: low-occupation status or low-education.

Primary Criterion. The study was included, if at least 70% of the sample was rated low-income-SES in the sense of earning a household income which is lower than 60% of the median equivalised income in the country where the study was conducted (for further information see section 1.2.2). Those thresholds are displayed in Appendix B.

If the mean or median household income of a sample was reported and...

... the corresponding standard deviation, then one standard deviation was added to the mean/median which complies with about 85% of the sample, assuming a normal distribution of the data. If this number was below the respective low-income-SES threshold (see Appendix B), the study was included.

... the maximum income, then the latter was added to the mean/median and divided by two which complies with about 75% of the sample, assuming a normal distribution of the data. If this number was below the respective low-income-SES threshold (see Appendix B), the study was included.

If no standard deviation or range was reported, the study was excluded, unless the mean/median was considerably below the threshold (i.e., if the doubled reported mean/median fell below the threshold).

With regard to the above mentioned inclusion, study authors needed to explicitly describe characteristics concerning the low-SES of the sample. Some exception were made, though, if certain indicators were given. For instance, if authors reported that all sample children came from centres in the US which serve low-income children only, the official poverty line was assumed and the study included. The official poverty line in the US lies below the here applied thresholds (see Appendix B): Even studies with samples whose income was 50% above the official poverty line, were included (the threshold is at 155% of the federal poverty line). Another example for inclusion was given, if the study was conducted in a low-income country according to the World Bank's definition[4] as well as assessing a sample which came from low-SES neighbourhoods.

There is also a group of indirect measures of low-income-SES: state aid programmes. Only those programmes are of interest here which are eligible exclusively or to a high percentage for families with low income. A widely spread example for intervention studies conducted in the United States is the quota of free- or reduced-price lunch children in the sample. Children are eligible for free-price lunch if their parents earn less then 130% of the federal poverty line in the United States, and up to 185% for reduced-price lunch eligibility, respectively[5]. Hence, a study with a sample containing 70% free- or reduced-price lunch children was included, since the mean of about 155% is below the low-income-SES threshold[6]. Other corresponding state aid programmes are for example: Aid to Families with Dependent Children (AFDC), Medicaid, MassHealth, Hartz IV, Arbeitslosengeld II, Supplemental Nutrition Assistance Program (SNAP)/Food Stamps, Temporary Assistance for Needy Families (TANF), or Alaska Temporary Assistance Program (ATAP).

Secondary Criteria. If no information about the income background of the sam-

[4] http://data.worldbank.org/about/country-classifications/country-and-lending-groups#Low_income

[5] Retrieved from United States Department of Agriculture website: http://www.fns.usda.gov/school-meals/income-eligibility-guidelines

[6] When taking, for example, the average of the federal poverty lines in the US from 2000 to 2011 and applying 155%, it results in an income of 15 315US$ per capita which lies below the low-income-SES threshold according to Appendix B for that time interval.

ple was reported, information concerning the occupational status or the education of the parents was employed as indirect criterion for measuring the primary criterion low-income-SES. Regarding low-occupation-SES, the study was included if at least 70% of the sample consisted of participants whose familial main occupation was rated as unskilled (e.g., porters, dustmen, labourer) or semi-skilled (e.g., postal workers, plant/machine operators), or if 70% of the sample lived in households where adults work less than 20% of their total work potential (e.g., part-time jobs, periods of unemployment). Exemplary scales that assess (among other social indices) occupational status are the Warner Scale (1960, occupations on level six and seven correspond to a low-occupation-SES), Duncan's Socio-economic index (O. Duncan, 1961, occpations up to level 15 correspond to a low-occupation-SES), or the occupation scale of the Hollingshead Scale (Hollingshead & Redlich, 1958, categories which include semi-skilled or unskilled occupations). If low-occupation-SES information were reported for mothers and fathers, the status of the father was considered as better representative for the familial main occupation status. That is due to the traditional assignment of roles, in other words, the mother as child-rearer and the father as the bread-earner, especially in older studies.

Concerning low-education-SES, the study was included if at least 70% of the sample consisted of families within which no parent achieved an educational level higher than ISCED 3 which is in industrialised countries equivalent to a lower secondary school degree or about 10 years of schooling.

Another secondary indicator for low-income-SES were certain socio-economic indices: The study was included if more than 70% of the sample were categorised as *low* on a socio-economic index, which is defined as scale that integrate several socio-economic measures. Thereby *low* had to correspond to the primary criterion. For instance, studies were included if more than 70% of the sample...

... were categorised as class IV on the Hollingshead Scale of 1958.

... stayed below a value of 25 on the revised Hollingshead Scale of 1975.

... stayed below a value of five on the Schichtindex by Lampert and Kroll

(2006).

... were rated higher than 62 on the McGuire-White Index of Social Status (McGuire & White, 1955).

Reasons for Exclusion

- **Clinical Sample:** The meta-analysis only included programmes with a preventive character (see section 2.1.1). Thus, studies with therapeutic contents, in the sense of treating mental and behavioural disorders, were excluded. This applied if the sample did consist of participants that were (clinically) diagnosed with a mental or behavioural disorder assigned by the DSM-IV, DSM-5, or ICD-10. Accordingly, programmes which were conducted in a mental hospital or a psychiatric clinic, were also excluded.

- **Basic Information of Low-SES:** Studies by authors who only reported vague or very simple sample descriptions without explicitly referring to the analysed sample, were excluded. That was among others the case for expressions like *low-income sample*, *sample from socially disadvantaged populations*, or *sample lives in social problem areas/neighbourhoods*.

- **Irrelevant State Aid Programmes:** Since the governmental aids are granted based on criteria other than exclusively or primarily low-income, studies with samples that are eligible for state aid programmes for which a low-income-SES cannot be presumed were excluded. Among those programmes are the following: Medicare, Medi-Cal, Oregon Health Plan, Soonercare, TennCare, BadgerCare, State Children's Health Insurance Program, Program for Women, Infants, and Children (WIC), Home Instruction for Parents and Preschool Youngsters (Hippy), Head Start (schools), Health Care Card (Australia), Title I.

- **No Low-SES Focus:** The study was excluded, if the authors unintentionally recruited a low-SES sample but actually focused on a different disadvantageous aspect. That was done based on the assumption that the intervention contents and aims are adapted to the need of a particular population and cannot simply be transferred to a low-SES population. This encloses interventions which are

primarily designed for children with special needs, such as developmentally delayed, physically challenged or chronically ill children, but also children of drug abusing mothers, incarcerated mothers, or mothers suffering from AIDS. As argued in the meta-analytic review of home visitation programmes by Sweet and Appelbaum (2004), such samples differ systematically from programmes targeting normally developing children. Further examples for samples that belong to this category are children living in foster care, AIDS-orphaned children, torture victims, children exposed to armed conflicts, traumatised children, or children who just had a brain injury.

2.1.3 Methodological Criterion

Randomised Controlled Design. Only those evaluation studies were included which randomly allocated participants to an intervention and a control group. This corresponds to level 5 on the Maryland-Scale (Farrington & Welsh, 2003). The randomisation procedure entails many advantages over non-randomised designs. The most important benefit is that the process minimises the plausibility of an observed effect being explained by other causes (be it measured or unknown confounding variables) than by the treatment (Shadish, Cook, & Campbell, 2002; Baytop, 2006). The unit being assigned to the conditions could be individuals (e.g., children or adolescents, parents, or teachers) or clusters which are aggregates of individuals (e.g., families, classrooms, schools, neighbourhoods, or communities). However, cluster-randomised controlled investigations usually have a smaller number of units and, thus, fewer such units are available for random assignment. Additionally, the individuals within a cluster are not independent from each other. Consequently, it cannot be assumed that the programme effect is unaffected of extraneous variables (Shadish et al., 2002). However, many preschool programmes take classes as unit of randomisation because it is the children's natural setting. Hence, cluster-randomisation was judged acceptable. In order to maintain the comparability of the groups, though, cluster-randomised controlled designs were only included if they assigned at least five clusters to each group. They were classified as level 4 quality on the Maryland-Scale (Farrington, 2003). When using random assignment, a pretest assessment is not essential (D. T. Campbell & Stanley, 1963). Thus, experimental designs with and

without pretest assessment were included. At least one intervention group had to be compared to one untreated (e.g., no treatment or wait-list control group), minimally treated (e.g., minimal treatment control group, attention control group), or traditionally treated control group (i.e., treatment as usual control group).

Minimal Sample Size. Studies were only included if the intervention and the control group each consisted of at least 25 participants. One reason is that by increasing the number of study participants, the effect size estimation becomes more accurate because of the smaller standard error (Carrasco, Volkmar, & Bloch, 2012; Viechtbauer, 2007). In addition, a minimum sample size is also favourable for assuming a normal distribution of the data and thereby being more independent of the study authors' analytical techniques. For achieving this, a sample size of 50 participants is recommended (Bortz & Schuster, 2010, p. 141). Consequently, only studies were included which allocated at least 25 participants each to the intervention and the control group. By setting the sample size to 25 per group, it can furthermore be assumed that a minimum of test power can be achieved (Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010). One last reason for a minimal sample size is the publication bias: Smaller studies with non-significant results are published less often than larger studies which are normally published independently of their results (Dickersin, 1990; Gilbody, Song, Eastwood, & Sutton, 2000; Thornton, 2000).

Reasons for Exclusion

- **Natural Fluctuation:** If more than 50% of the original sample altered during the programme conduct due to natural fluctuation (e.g., students moving to different neighbourhoods and changing the school) and the newcomers where not assigned randomly to the groups, it was interpreted as a threat to the randomisation process and, hence, the study was excluded. No exclusion was done with regard to sample attrition.

- **Minimal Group Difference:** If the intervention and the control group only slightly differed regarding minor characteristics, the study was excluded. One example for that would be the sex of the programme executor, the usage of different media types (e.g., computer versus print, technology-enriched versus

traditional classroom), the length of the intervention (e.g., long versus short duration, half- versus full-day care), or the modulation of class composition (e.g., small versus large class size, group versus private lessons).

2.1.4 Outcome Criterion

Studies were only considered for inclusion if at least one dependent variable was measured within three months after the intervention ended and, additionally, if at least one dependent variable could be assigned to the primary outcome category *child outcomes*. Child outcomes was divided into two outcome constructs, that are basic and school development (see section 1.4.3). Regarding those two constructs, the children or adolescents themselves had to be the subject of assessment (i.e., the parents could be the informant). A detailed description of the primary outcome category with the corresponding constructs, measurement examples, and typical questionnaires or rating scales, is displayed in Table 3. For more details see Appendix H.

Further outcomes were included if they fit one of the two secondary outcome categories. If an assessment method (e.g., *Home Observation for Measurement of the Environment*) was chosen and included relevant outcomes (i.e., learning materials at home or activities offered by the parents) among irrelevant ones (e.g., modelling of social maturity), the outcome was included.

As for the primary outcome categories, Table 4 shows corresponding information for the two secondary outcome categories which were not further divided into constructs.

Table 3

Primary Outcome Category: Child Outcomes

Construct	Examples	Q/R
Basic development	Cognitive development outcomes (i.e., intelligence, reasoning, attention, comprehension (logic), phonological and morphological awareness, creativity, problem solving) and general development outcomes (i.e., general (infant) development, motor skills, exploratory behaviour, developmental quotient, school readiness)	BSID-MDI, BSID-PDI, DDST-"FM", IBS, RPM, Wechsler
School development	Academic achievement outcomes (i.e., reading/writing skills, repetition, spelling, mathematics or science achievement/skills, school grades) and educational development outcomes (i.e., learning problems, academic motivation/self-concept, drop out, discipline referrals, winning award, job satisfaction, income)	ATSQ, HHS, NSB, PASS, PPVT, WJ-"WI"

Notes. Q/R = Questionnaires or Rating Scales, BSID-MDI = subscale "Mental Development Index" of Bayley Scales of Infant Development, BSID-PDI = subscale "Psychomotor Development Index" of Bayley Scales of Infant Development, DDST-"FM" = subscale "Fine motor" of Denver Developmental Screening Test, IBS = Infant Behavior Scale, RPM = Ravens Progressive Matrices, Wechsler = Wechsler Intelligence Scale for Children, ATSQ = Attitudes Toward School Survey Questionnaire, NSB = Number Sense Brief Total, HHS = Hollingshead Score, PASS = Perception of Ability Scale for Students, PPVT = Peabody Picture Vocabulary Test, WJ-"WI" = subscale "Word Identification" of Woodcock Johnson.

Table 4

Secondary Outcome Categories: Parental and School Support

Category	Examples	Q/R
Parental support	Parental beliefs/wishes/expectations about child's educational career, parental reading ability, reading aloud, verbal responsiveness, giving privileges for good grades, telling stories, attending school meetings, provision of learning material or experiences	POI, HOME, PCERA
School support	teachers sensitivity, teacher-student interaction/closeness, application of certain teaching methods (simplification, student-centred practices), enjoyment of teaching, classroom climate, feeling of safety	CLASS, ECERS-R, IQA

Notes. Q/R = Questionnaires or Rating Scales, CLASS = Classroom Assessment Scoring System, ECERS-R = Early Childhood Environment Rating Scale, IQA = Instructional Quality Assessment, POI = Parent Opinion Inventory, HOME = Home Observation for Measurement of the Environment, PCERA = Parent Child Early Relational Assessment.

Outcomes not Considered for Analysis

- **Socio-emotional Outcomes:** Outcomes which measured the socio-emotional development of the children or adolescents were not integrated in this analysis. Such outcomes are, for example, internalising behaviour (e.g., depression, anxiety, emotion regulation, withdrawal, tension/relaxation), externalising behaviour (e.g., conduct problems, hyperactivity, delinquent behaviour, incarceration, deviance), self-related cognition (e.g., self-esteem, self-concept, self-confidence, coping), social competence (e.g., social interaction, empathy, apathy, expressiveness, adapting skills), sexual parameter (e.g., risky sexual activities, number of pregnancies or births), or substance abuse parameter (e.g., substance abuse, attitude toward substance consumption). This also applied to teacher-related socio-emotional outcomes, such as dealing with mobbing, bullying, or peer discrimination.
- **Cost-effectiveness Outcomes:** Variables concerning the economic evaluation of an intervention were not included into the analysis.

- **Upbringing Outcomes:** Outcomes which measured parental attitudes, knowledge, or behaviour in terms of supportive upbringing strategies (e.g., self-efficacy, authoritarian educational beliefs, rejection of child-reading role, corporal punishment, neglect, victimisation, abuse, parenting skills in general) were excluded from the analysis.

Reasons for Exclusion

- **Secondary Outcomes Measured Only:** If only secondary outcomes were assessed and no primary outcome, the study was excluded.

2.1.5 Effect Size Criterion

The studies needed to present results in such a way that allowed a reliable calculation or estimation of effect sizes that directly (assessing the child) or indirectly (assessing with the child associated third party) referred to a sample described under section 2.1.2. At least one direct measure had to be assessed. Different types of effect sizes where integrated if they lead to comparable results and made an integration between the included studies possible (see section 2.4).

Reason for Exclusion

- **Polarity Unclear:** If the direction of the effect was unclear (i.e., higher values stand for better development or vice versa), the study was excluded (see section 2.4.7).

2.1.6 Duplicates Criterion

If several studies exist which analyse the same content and sample, one main study was determined and only this study was included in the analysis. This prevented data doubling and, thus, the distortion of the results by authors who are more affine to publications. The other studies are referred to as duplicates, differentiating between to kinds: supplemental and partial/complete duplicates. A supplemental duplicate contains valuable information that were not published in the main study, such as further outcome assessments or follow-up assessments. This information was used

for the coding process. A partial or complete duplicate has in core the same content as the main study with some minor variations (e.g., slightly different number of participants or published in a different journal). Duplicates were summarised under the respective main study.

2.2 Study Search

In order to find all relevant studies which contribute to answering the research questions, a thorough literature search was conducted. Especially an extensive worldwide search in different online databases was considered important to take into account the intended generalisability of the findings. The first step was to scan the included studies of other syntheses reports pertinent in this research field (see section 2.2.1). Subsequently, the main search began (see section 2.2.2): The resulting initial primary studies were used to facilitate the development of a search strategy for the online databases by scanning the title, abstract, and keywords for possible search terms. After the online search was finished, the studies were scanned for inclusion. In a final step, all references of the included primary studies were reviewed for eligibility.

The search has been conducted between January 2013 and November 2014. All studies published until December 2013 were included.

2.2.1 Initial Search

As initiation strategy for the search, meta-analyses and reviews which did already synthesise studies of a similar topic within the last 15 years were gathered (see Appendix C). The screening of the studies which those meta-analyses included led to a first pool of primary studies. The title, abstract, and keywords of those studies were scanned for possible search terms. Those were used to built the base for developing the main search strategies for the online databases.

2.2.2 Main Search

In the area of psychology, sociology and pedagogy, the Thüringer Universitäts- und Landesbibliothek provides access to 185 data bases. However, most of them are

dictionaries, encyclopaedias, factual databases, inventory listings, or bibliographical databases. This leaves a number of 13 databases which contain thematically pertinent articles, access sufficiently distinctive sources, have at least basic search options, and provide a possibility to export the found citations.

The databases can be searched directly or can be accessed via interfaces. Since searching through interfaces provides the option to scan several databases simultaneously (e.g., ERIC and Medline can be entered via the interface OvidSP), this technique was employed, if possible. As a common base, the eligibility criteria were used for the content and the structure of the search. Then, the search strategy for each database or interface was developed using three central functions: Search fields to adapt the content, Boolean operators to adapt the structure and replacement options to refine the strategy. Appendix D presents the different interfaces and the therein searched databases, as well as the search fields and features used for the search strategy. Not depicted in Appendix D are the search options that all the selected databases have in common, namely Boolean Operators, quotation marks and parentheses. The search fields, Boolean Operators and replacement options will be explained in more detail in this next section, finalised by a paragraph about evaluating a search strategy.

Search Fields: A Search's Content. Adjusting the search content is an act of balancing between a too broad search and, thus, finding too many irrelevant articles, and searching too narrowly which could results in missing relevant articles. For the thesis, the commonly used search field *title* and/or *abstract* proved successful for keeping the balance. Additionally, the search fields *keyword* and *subject* were integrated. Since those terms are used inconsistently in the literature and in the databases, they are explained in detail in Appendix E. Apart from those four search fields, other interface-specific ones were used. Appendix D gives an overview of the search fields with their technical notation for each interface in brackets.

Boolean Operators: A Search's Structure. In order to develop a complex search strategy, search options are required which allow the determination of the relationship between the search terms and, thus, enable structuring the search concepts: Boolean operators. They trace back into the 19th century to the English mathematician George Boole. He aimed at describing philosophical interrelations by using mathematical expressions. He used the logical values TRUE and FALSE and the

functions AND, OR, and, NOT and therewith laid the basis for all computer and programming languages until today (Breitsameter, Hillenbrand, & Burgwitz, 2009). For systematically searching databases, the three functions AND, OR, and NOT are essential. Since they are used in all for this meta-analysis relevant databases, no adjustment was needed concerning the basic structure of the search. For all interfaces, the structure consisted of four main blocks corresponding to the central eligibility criteria (see Table 5, block 1 to 4). Each study had to investigate the effectiveness of a psychosocial or educational intervention (block 1) which aims at preventing the negative consequences of low-SES (block 2) for children and adolescents (block 3), by using a controlled study design (block 4). Within each block, the different search terms paraphrasing the respective criterion were connected applying the OR-operator. Between the blocks the AND-operator was used, searching for studies that fulfilled each of the four criteria. If suitable, database-specific additional blocks were added to refine the search. Regarding the strategy for OvidSP, for instance, a block for pharmacological therapies was drawn off the main strategy by using the NOT-operator.

Table 5

Structure of the Online Search Strategy

Block 1 PS&E intervention	prevention **OR** intervention **OR** therapy **OR** programme **OR** training **OR** tutoring **OR** ...
	AND
Block 2 Sample of low-SES	poverty **OR** economically disadvantaged **OR** low socio-economic status **OR** low income **OR** ...
	AND
Block 3 Children and adolescents	children **OR** youths **OR** teens **OR** adolescents **OR** baby **OR** pupil **OR** minor **OR** ...
	AND
Block 4 Controlled design	control group **OR** comparison group **OR** controlled trial **OR** randomised-controlled trial **OR** ...

Replacement Options: Refining the Search. To refine a search strategy and, thus, to achieve a better hit rate, it is possible to work with search features: quotation marks (enable *phase searching*), parentheses (prioritising and grouping of search terms), proximity operators (fixing the distance between search terms), truncations (searching for variations of a term), and wildcards (place holder function). It is important to mention that quotation marks and parentheses are universal features, just like Boolean operators, whereas proximity operators, truncations and wildcards may differ between the used databases (see Appendix D). For more details on how to apply the replacement options, see Appendix E.

Indicator for the Quality of the Search Strategy. No objective or commonly used criteria exist for evaluating the quality of a search strategy. This can be explained by the many available options for designing searches and the resulting infinite possibilities for a search strategy. They may differ strongly but each can still yield satisfying results. The time and economic framework also need to be considered in this context. Since the search strategies of the thesis at hand were developed within the project PrAMi (see beginning of section 2), the search was designed much wider than necessary for the topic. This has the advantage to maybe finding studies which might not have used the typical nomenclature for describing the study content. For evaluating the efficiency of the strategies regarding this meta-analysis, the following criteria were applied:

- **Rough precision:** A first impression concerning the effectiveness was received by test scanning the titles and abstracts of 50 randomly selected articles found by the corresponding search and analysing how many studies would be included. As a thumb rule, if the ratio of included to found studies exceeded 10%, the search was judged satisfying. If that was not the case, the search was adapted.

- **Hits found:** Another approach was to check if the studies already included by scanning similar meta-analyses (see section 2.2.1) could be found by the respective search. If they were not retrieved, it was analysed, which search terms were missing and the search correspondingly improved.

- **Non-hits are found:** This strategy aims at dismissing non-hits by analysing

which search terms resulted in finding studies which are not of relevance and changing them.

The final search strategies for the different interfaces are presented in Appendix F. With those strategies more than 6 000 articles were found. In a first step, all studies were scanned by title, abstract, and keywords for eligibility. If the information needed for a decision were not available, the full text was requested. If it was uncertain whether to include the study or not, it was discussed in the weekly meetings of the project PrAMi. Regarding all uncertainties and coding problems, a solution was found based on common consensus.

In a last step, the references of all included studies were scanned for eligibility.

2.2.3 Search Results

The process of the different search steps is summarised in Figure 3. Through scanning online databases, 6 029 scientific reports could be identified after removing exact duplicates. By scanning scientific overviews[7] or the references of included primary studies, and by chance findings[8], another 851 scientific reports were identified, yielding to a total number of 6 880 scientific reports. After scanning the title and the abstracts, 4 720 articles were excluded. For the remaining 2 160 studies, the full texts were requested. If articles were not obtainable by the ordinary ways, the articles' authors were contacted. As a consequence of those efforts, only 50 studies remained unavailable (less than 3% of the potentially relevant scientific reports). Furthermore, if applying the ratio of included studies and studies scanned for eligibility to the 50 unavailable reports, only three studies were theoretically missed[9]. It is very likely that those would have been duplicates of already included studies.

[7] A scientific overview includes meta-analyses, reviews, summaries, overviews, summarising books and the like.

[8] During the search process, studies were also found by chance. That was the case, for instance, when looking for a full text on the homepage of a ministry and thereby scanning the other programmes offered there. If one of those fit the eligibility criteria, it was included in the analysis. Another example concerns studies which were found by the PrAMi-migration search strategy (see beginning of section 2).

[9] $= \frac{109*100\%}{2110} = \frac{5.2\%*50}{100\%} = 2.6$

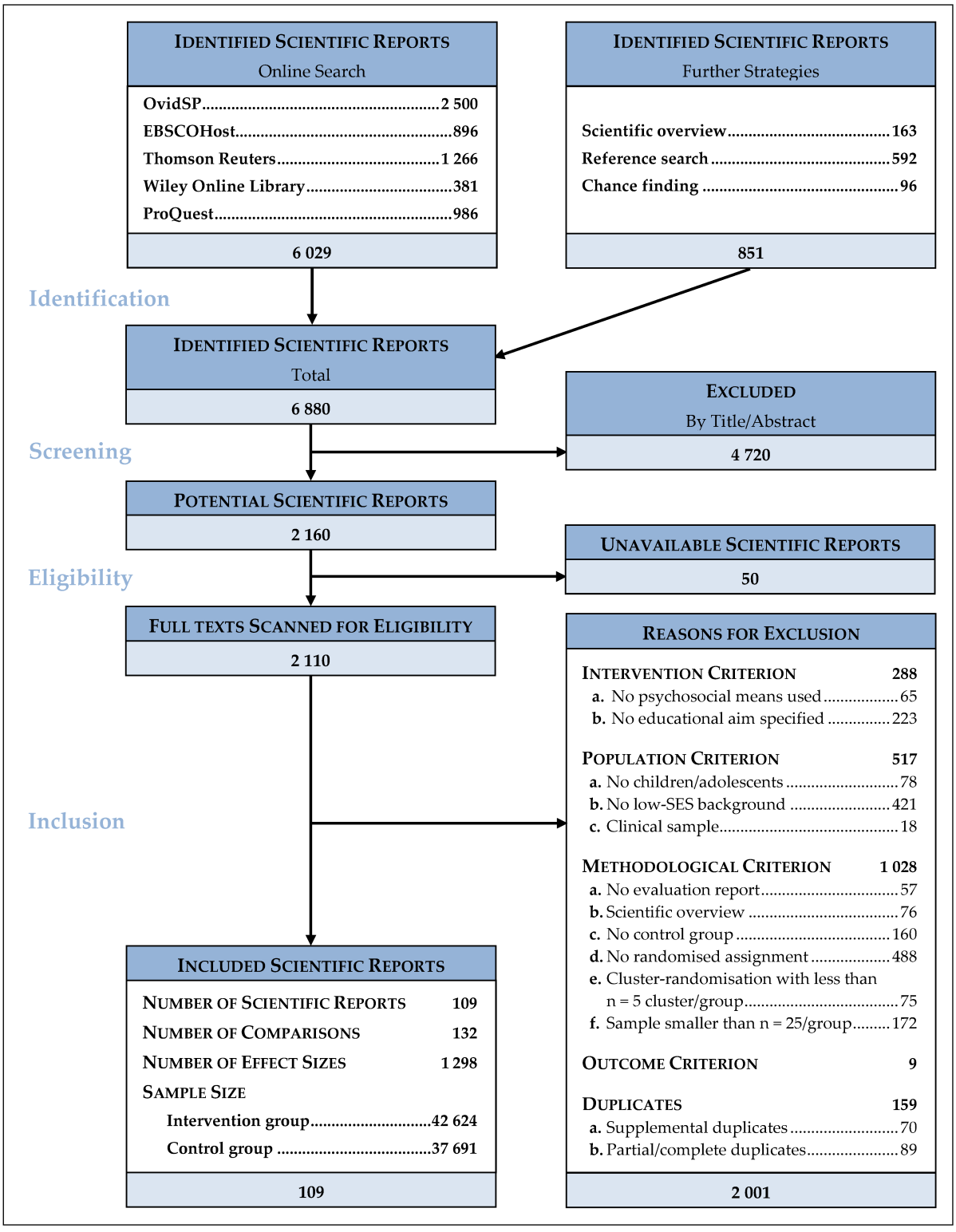
Each accessible full text was then scanned for eligibility which lead to the exclusion of 2 001 scientific reports.

The reasons for exclusion are grouped according to the eligibility criteria. Most of the reports (51%) were excluded because of methodical reasons, followed by the population criterion with 26%. About 15% of the studies had to be taken out of the pool for not fitting the intervention or outcome criterion. Of the remaining 268 reports, 159 were identified as duplicates leaving a primary study pool of 109. Those comprised of 132 comparisons (the term will be explained in the next section) with 1 298 effect sizes coded, hence, about 10 effect sizes per comparison. The overall sample size was 80 315.

All included studies are displayed in Appendix G.

Figure 3

Flow Chart Showing the Different Stages of the Study Search



2.3 Coding Process

After searching and selecting the primary studies, the necessary information for answering the research questions need to be retrieved. For this purpose, a coding handbook was developed together with a master data sheet. Moreover, a code training was conducted and, finally, the studies were coded.

2.3.1 Coding Handbook

For systematically collecting information from the primary studies, a coding handbook was developed within the project PrAMi. It contains the variables of interest for answering the research questions. The first ideas for interesting and relevant variables were inspired by meta-analyses and reviews pertinent in this research field. Hence, the first version of the coding handbook was developed already around the time when the study search started (January 2013). Over the next months, individual test studies were coded to check the coding feasibility of each variable and to ensure that all relevant variables were included. At the beginning of September 2013 the handbook was so well revised with regards to the content that it was handed to Dr. Marrilyn Hooley[10] who lectured it. With the lectured version, a final test coding was conducted at the end of September 2013. After some last changes, the training of the coders began in October 2013. During the next two months, 10 studies were coded aiming at an average inter-rater reliability (Cohen's Kappa) of more than $\kappa = .80$ between the scientific assistants (Angelika Schulz, Sandra Renas, Katrin Konold) and the Ph.D. students (the author, Dipl.-Psych. Louisa Arnold, Dipl.-Psych. Sebastian Schulz). After some last refinements of the coding handbook, the final version was obtained and the main coding period started in December 2013. It lasted until May 2014. For this meta-analysis the coding handbook was adapted regarding, for instance, the outcome structure (see Appendix H).

The coding handbook itself is hierarchically organised in three levels: publication, comparison, and effect size level. By assigning an identification number (ID) to each

[10] Dr. Hooley is a lecturer at the Faculty of Health in the Department of School Psychology of the Deakin University in Melborune, Australia.

publication on publication level (the term *publication* is here applied synonymously for *study* or *scientific report*) and transferring the same to the other levels, the data was connected across all levels. On each subordinate level a new digit was added, whereby each effect size has its unique ID and can be traced back to its corresponding comparison and publication.

The structure for each publication is visualised clearly by the master data sheet (see section 2.3.2).

Publication Level. On publication level general information were collected, such as the author(s) or the title of the publication and the publication year. Also listed here were the publication format (e.g., doctoral thesis or journal article) and language, as well as the country where the study was implemented. The last part concerned collecting information about potential duplicates which were used for coding the main study.

Comparison Level. The term *comparison* refers to the contrast between an intervention and a control group. It will be the main analytical level in this dissertation, thus, the majority of the variables were assessed here.

On comparison level, information about characteristics of the intervention and the sample were gathered, as well as information for evaluating the validity.

Regarding information about the intervention, first basic programme information were gathered, such as its aim(s), content(s), and the programme type. *Programme type* had a special selective function: If a programme addressed, for example, the parents and the child to the same degree, *child training programme* and *parent training* was checked. For the remaining variables on comparison level information of both addressees were considered for coding. Disadvantageous about this strategy is that all following variables had to be averaged: If the parental component was structured *low* and the child component *high*, the variable was coded *medium*. Hence, there is a great loss of information. Therefore, it was tried to avoid this by opting for one *primary* programme type: If the programme had some parental training but focused on working with the child, the parental part was neglected for the rest of the variables on comparison level. However, the information regarding the parent training component was not completely lost, since it could be coded in the variable *Programme content*. The alternative option would have been to code all variables for the parental

component and the child component, separately. However, this would have enlarged the coding handbook to an uneconomic extent.

Following the basic information, a part which contained rather technical information about the intervention was integrated, assessing, for example, the theoretical foundation of the programme, the implemented methods for conducting it, the use of a manual, its degree to adapt programme methods, and the grade to which the participants actively took part in the programme. Several variables were created also for rating the quality of programme implementation, like the amount of training and supervision the executor received, or whether conducting the programme was observed and documented.

The next part contained variables which assess the intensity of the intervention (e.g., length or the number of sessions conducted), geographical information (e.g., whether the area where the intervention was implemented was deprived or which setting was chosen), and information about the executor and developer (e.g., the executors profession or the relationship to people associated with the publication).

Following that, information were collected regarding the sample, such as a closer description of its socio-economic situation and its ethnic composition, but also the children's age and gender, as well as additional risk factors. The recruitment, attrition, and attendance rate were also allocated here.

In a last section, several variables were assessed for evaluating the validity. Those were based upon the criteria by Cook and Campbell (1979b) and included, for instance, the type of control group, the study design, or the group equivalence at baseline.

Effect Size Level. On effect size level information concerning the effectiveness measures can be found. First of all, the outcomes were hierarchically systematised in categories and constructs. Subsequently, variables regarding the assessment of the outcome(s) were placed, such as the kind of measurement used, the person who assessed the data, and the person who has been assessed. Then, all information necessary for calculating the effect sizes were gathered (allowing up to two follow-ups), as well as variables for validity evaluation (e.g., the degree of standardisation or the level of impact).

Challenges when Developing a Handbook. There are some challenges when cre-

ating a coding handbook that cannot be solved but should be kept in mind. The issue complexity versus economy for example: The more complex a coding handbook is designed, the more probable it is that it grasps the different kinds of interventions (e.g., child-focused play-like group intervention at school versus information-based individual parent training at home). However, a complex handbook also increases the time needed to code each study and, further, makes is harder to synthesise the results or obtaining a sufficient number of cases in each cell for analysing moderating influences on the average effect.

Another challenge concerns the contrast between accuracy and reporting. It is preferable to aim at creating variables which gather precise information concerning the publication in order to describe it more accurately. Nevertheless, the more specific the information needed is, the fewer publications might have reported it, leaving many missings and making it hard to analyse the variable at all.

2.3.2 Master Data Sheet

Complementing the handbook, a *master data sheet* was developed where the hierarchical structure of each publication was noted. Using the master data sheet allowed the coders to easily enter coding discussions by having an immediate overview of studies they didn't read. In Appendix I, the master data sheet is exemplarily filled out for the primary study of Cronan, Cruz, & Arriaga (1996): On the first page central information were noted, such as the ID of the main study and the name of the coder, as well as check boxes regarding the fulfilment of the eligibility criteria. On the second page, the structure of the publication was noted. In case of Cronan, Cruz, & Arriaga (1996), the publication contained one study with two intervention groups and one control group, resulting in two comparisons. The two comparisons have the same outcome structure: Eight outcomes concerning the home environment were measured once one week after the intervention ended.

2.3.3 Coding Quality and Challenges

Two central demand which were aspired, concerned a high transparency and objectivity of the coding process. Hence, different measures were taken into account in

order to accomplish that. First of all, the efforts put in and the time spent on developing the coding handbook (see chapter 2.3.1) resulted in precisely defined variables and concretely formulated coding guidelines (see Appendix J). The latter were updated weekly during the coding process. Second, using the master data sheet had not only the advantage of better breaking down complex study structures into little chunks, but also made it possible to easily compare and by that discuss the studies. Another measure taken was conducting an intensive coder training in order to maximise the coding agreement and the coding routine. Only after reaching an interrater reliability of more than $\kappa = .80$ (Cohen, 1960), the coders could code independently. However they were still involved in a weekly discussion group which was established for all coders to discuss uncertainties. For those meetings the minutes where kept, so that it was possible to review the coding decisions and, by that, establishing a better consistency over time. All those measures maximised the transparency, objectivity, and comparability of the coding process.

Of the 109 included scientific reports the author of the dissertation coded 84. The remaining ones were coded by Sebastian Schulz (n_P = 24) and Louisa Arnold (n_P = 1).

2.4 Calculating Effect Sizes

In this meta-analysis one effect size always represents a group difference. In other words, the effect between an intervention and a control group after the programme was conducted. One challenge lied in the variety of outcomes measured and operationalisations used in the different studies: Usually a study author measured several outcomes and sometimes with different tests. In order to integrate the various effect size estimates, it was important to standardise them. Thus, each relevant outcome variable was transformed into a comparable effect size *d* and then optimised by the correction factor *J* suggested by Hedges (1981). For the effect size estimations, pertinent meta-analytic books (Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Cooper et

al., 2009) and relevant tools, such as the Effect Size Calculator by Dr. Wilson[11] or the Statistics Calculator by Dr. Soper [12], were consulted.

The basic guideline for obtaining the effects was to rather include a highly estimated effect size than loosing information regarding the results by excluding it. Nevertheless, the more accurate information was always preferred over a less accurate one. The following list provides the applied hierarchy (from most precise calculation to highly estimated):

1. **Effect Sizes Based on Standardised Mean Differences**: Availability of means, standard deviations, and sample sizes of the intervention and control group (before the beginning of the programme (pretest) and) after the programme ended (posttest).

2. **Reported Effect Size:** Effect sizes calculated by the study authors.

3. **Back Calculation by Statistical Parameters:** Availability of statistical parameters, like *t* test, *F* test, correlation coefficient *r*, or Odds Ratio plus the sample sizes in both groups at posttest.

4. **Estimated Effect Sizes:** Report of non-significance or literal information only.

2.4.1 Effect Sizes Based on Standardised Mean Differences

If authors reported the means, standard deviations, and sample sizes at posttest, a standardised mean difference could be calculated based on the formula by Lipsey and Wilson (2001): Subtracting the mean of the intervention ($\overline{X}_{IG}$) and the control group ($\overline{X}_{CG}$) and dividing them by their pooled standard deviation (s_{pooled}):

[11] David B. Wilson is professor at the Department for Criminology, Law and Society at the George Mason University in Fairfax (VA), United States. The Effect Size Calculator can be found under: http://campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD1.php (retrieved on 25.09.2015)

[12] Dr. Daniel S. Soper is an associate professor at the Department of Information Systems and Decision Sciences at the California State University in Fullerton (CA), United States. The Statistics Calculator can be found under: http://www.danielsoper.com/statcalc3/default.aspx (retrieved on 25.09.2015)

$$d = \frac{\overline{X}_{IG} - \overline{X}_{CG}}{s_{pooled}} \tag{1}$$

The pooled standard deviation at posttest is then,

$$s_{pooled} = \sqrt{\frac{(n_{IG} - 1) * s_{IG}^2 + (n_{CG} - 1) * s_{CG}^2}{n_{IG} + n_{CG} - 2}} \tag{2}$$

All parameters refer to the posttest time-point, where n_{IG} is the sample size of the intervention group, n_{CG} is the sample size of the control group, s_{IG} is the standard deviation for the intervention group, and s_{CG} is the standard deviation for the control group. The variance of the standardised mean difference is:

$$V_d = \frac{n_{IG} + n_{CG}}{n_{IG} * n_{CG}} + \frac{d^2}{2(n_{IG} + n_{CG})} \tag{3}$$

If pretest data was available, they were included in the effect size calculation to consider possible differences between the groups which might have occurred despite random assignment. Based on Morris' formula (2008), the mean difference between the intervention and control group at pretest ($\overline{X}_{IGpre} - \overline{X}_{CGpre}$) was subtracted from the mean difference between the groups at posttest ($\overline{X}_{IGpost} - \overline{X}_{CGpost}$) and divided by the pooled standard deviation was pretest ($s_{pooled,pre}$):

$$d = \frac{\overline{X}_{IGpost} - \overline{X}_{CGpost} - (\overline{X}_{IGpre} - \overline{X}_{CGpre})}{s_{pooled,pre}} \tag{4}$$

Morris recommends using the pooled standard deviation at pretest and not at posttest because it results in a more accurate effect size estimation. One reason is that the pooled standard deviation at pretest is unbiased concerning the intervention effect. The pooled pretest standard deviation is calculated as follows:

$$s_{pooled,pre} = \sqrt{\frac{(n_{IGpre} - 1) * s_{IGpre}^2 + (n_{CGpre} - 1) * s_{CGpre}^2}{n_{IGpre} + n_{CGpre} - 2}} \tag{5}$$

where n_{IGpre} and n_{CGpre} are the sample sizes of the intervention group and control group, respectively, and s_{IGpre} as well as s_{CGpre} are the corresponding standard deviations. All parameters are measured at pretest. The variance was calculated according to formula 3.

2.4.2 Reported Effect Size

If the reported information where not sufficient to calculate a standardised mean difference but the authors reported an effect size, it was included in the meta-analysis. Usually, the authors also reported the employed formula or the exact effect size parameter with the corresponding reference, which fitted the here used formulas. This strengthened the assumption that reported effect sizes can be judged as reliable parameters, similar to self-calculated standardised mean differences.

2.4.3 Back Calculation by Statistical Parameters

Effect sizes can also be estimated with relative high precision if only statistical parameter and sample sizes are reported. For example, in case of a reported t-value or F-value from univariate analysis of variance (ANOVA) plus sample sizes in both groups, *d* was calculated as follows:

$$d = t * \sqrt{\frac{2}{N_{harmonic}}};\ N_{harmonic} = \frac{2 * n_{IG} * n_{CG}}{n_{IG} + n_{CG}} \tag{6}$$

$$d = \sqrt{\frac{F * (n_{IG} + n_{CG})}{n_{IG} * n_{CG}}} \tag{7}$$

If the exact p value and the underlying distribution was reported, the statistical parameter was calculated and the formulas above applied. If a simple linear regression with group (intervention and control group) as independent variable and the outcome as dependent variable was reported, the regression coefficient was interpreted as correlation and the effect sizes calculated as follows:

$$d = \frac{2r}{\sqrt{1 - r^2}} \tag{8}$$

The variance for effect sizes based on statistical parameters (formulas 6 and 7) or regression coefficients (formula 8) were calculated based on the formula regarding standardised mean differences (formula 3). For effect size calculation based on binary data, the following formulas where applied:

$$d = logOR * \frac{\sqrt{3}}{\pi};\ logOR = \frac{a * d}{b * c} \tag{9}$$

whereas *a* and *c* represent the occurrence of the event in the intervention and control group, respectively, and *b* and *d* stand for the occurrence of the complementary event in the intervention or the control group, respectively. The variance of *d* is then,

$$V_d = V_{logOR} * \frac{\sqrt{3}}{\pi^2};\ V_{logOR} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \tag{10}$$

2.4.4 Estimated Effect Sizes

To the group of estimated effect sizes belong those with a considerate level of uncertainty. If effects were only described as *non-significant results*, for instance, the effect size was fixed to $d = 0$. If an outcome was reported as being significant on the 5% level, a *p* value of .05 was used and a *t* distribution of the data assumed. That allowed a back calculation as described in section 2.4.3. Those strategies are conservative in the sense that, if affected, the mean effect size would be underestimated. However, that was preferred over loosing information. Also judged as considerably estimated were effects which were confounded. That was, for example, the case for multiple regressions, where the unstandardised regression coefficient (if reported with standard error) or the *t* value of the comparison between the intervention and the control group served as best possible estimator (if no other information was available). In a similar way, the *F* value of reported multivariate analysis of variance (MANOVAs) or analysis of covariance (ANCOVAs) which included an intervention versus control group variable was used for effect size estimation. If the author only literally described the effect of an intervention compared to a control group concerning a certain outcome with *small*, *medium*, or *high* (or the like), the effect size was estimated with $d = 0.20$, 0.50, or 0.80, respectively, according to Cohen's evaluation (1988). If the only source

of information was in form of a graph, the coder estimated the effect. All estimated effect sizes were discussed in the weekly group meeting which all coders attended (see section 2.3.3). For each discussed effect size a consensus decision concerning the estimation could be found.

2.4.5 Correction Factor J

According to Hedges (1981) the effect sizes calculated under sections 2.4.1 to 2.4.4 result in an overestimation if the sample sizes are small ($N < 10$). Therefore, Hedges introduced a correction factor J,

$$J = 1 - \left(\frac{3}{4 * df - 1}\right);\ df = n_{IG} + n_{IG} - 2 \tag{11}$$

which is multiplied with the effect size d resulting in an unbiased effect size d_*:

$$d_* = d * J \tag{12}$$

2.4.6 Outliers

Subsequently to the calculations, it was checked if there were outliers among the effect sizes, defined as effect sizes which are lower/higher than $d_* = \pm 3$. That was the case for seven effect sizes. All of them were based on unrealistically low standard deviations at pretest time-point. By applying the posttest time-point standard deviation instead, all outliers could be corrected and, thus, no effect size was excluded.

2.4.7 Polarity

In order to compare the different effect sizes, it is necessary that they signify the same regarding their interpretation. Here, a positive value stands for a positive effect of the intervention compared to the control group (positive polarity or in favour of the intervention group), whereas a negative value represents a positive effect of the control compared to the intervention groups (negative polarity or in favour of the

control group). If the direction of an effect was unclear, the effect size was excluded from the analysis, since it was not possible to interpret the effect.

2.5 Aggregating Data

As mentioned at the beginning of section 2.4, it was possible and typical that several different outcomes where measured within one comparison, applying even different tests for the same outcome. The most common example was measuring language by different tests and on several time intervals. Since the main analytical level was the comparison level (see section 2.3.1), it was necessary to summarise the effect sizes within each comparison. This follows a comparable logic as the one applied in section 2.1.6: Preventing the distortion of the results by authors who are more affine to assessing many outcomes, compared to those who assess few. By averaging the strongly differing numbers of effect sizes, the data was aggregated to comparison level and, hence, each comparison weighted equally.

2.5.1 Variables of Interest

Aggregating the data can be realised in different ways depending on variables which might influence the effectiveness of the intervention. Given the earlier mentioned diversity in outcome assessment, it is reasonable to expect variations concerning the effectiveness depending on different outcome categories or constructs. Furthermore, it is expected that the intervention effectiveness changes over time.

Since the focus of the thesis lies on the educational development of children and adolescents, they were to be analysed more differentiated. Thus, the first variable of interest was *outcome construct*, containing the two under section 2.1.4 illustrated outcome constructs: basic and school development. Then, a rougher classification variable *outcome category* was established with three categories: child outcomes, parental support, and school support. The final variable of interest was the time interval between the end of the intervention and the assessment time-point. Three time intervals were defined for this purpose. The first included all outcomes assessed after the intervention ended until three months post and was correspondingly named short-term effect. The second time interval contained mid-term outcomes assessed later than

three months after the intervention ended up to maximally 12 months post. Long-term effects comprised the last time interval with effect sizes measured later than 12 months post intervention.

For integrating the data to comparison level, the effect sizes were averaged per comparison within each outcome construct for each time interval as well as within each outcome category for each time interval, resulting in 15 averaged effect sizes (two outcome constructs x three time intervals + three outcome categories x three time intervals) that each comparison could maximally have. Additionally, within each comparison, each outcome construct and each outcome category were averaged across all time intervals, adding another five averaged effect sizes (two outcome constructs + three outcome categories). Finally, the outcome categories were averaged within each time interval per comparison, resulting in additional three averaged effect sizes (one for each time interval). Altogether, each comparison could maximally have 23 averaged effect sizes (15+5+3).

2.5.2 Aggregation Strategy

Before aggregating the data across all comparisons, a decision concerning the underlying model needed to be taken. The two possible models are the fixed-effect model (FEM) and the random-effects model (REM, Borenstein, Hedges, Higgins, & Rothstein, 2009). The models differ in the assumption concerning the variance of the effect sizes across all comparisons. The FEM assumes that a true overall effect size exists and, thus, variability between effects across comparisons is attributed to sampling errors (i.e., random variance within each study or within-study variance; WSV). In contrast to that, a second type of variance is included in the REM, referred to as between-study variance (BSV). The model assumes that the differences of the effect sizes between the comparisons are true differences in the sense that other variables are expected to influence the true overall effect size. In this thesis, the existence of true variances between the comparisons was expected due to the high variability between the comparisons concerning, for instance, the type of intervention, the sample, or the circumstances of programme conduct. Thus, for the aggregation of the data, the REM was presumed. The assumption of an REM, moreover, allowed to analyse the influence of moderators on the effectiveness which was one of the major foci of

this thesis.

As already mentioned in section 2.1.3, effect sizes become more accurate the higher the number of assessed study participants is (Carrasco et al., 2012; Viechtbauer, 2007). In order to consider this when calculating the overall effect size $\overline{d_*}$, the effect sizes d_{*i} where weighted by their inverse variance w_i. This is considered to be a more accurate approach than weighting by their sample size (Cooper et al., 2009). The integration method is based on Hedges and Olkin (1985):

$$\overline{d_*} = \frac{\sum_{i=1}^{k} w_i * d_{*i}}{\sum_{i=1}^{k} w_i};\ w_i = \frac{1}{v + SE_i^2} \tag{13}$$

The denominator of the weight w_i in formula 13 contains the assumptions of the REM, whereas v_i is the BSV and SE_i^2 is the squared standard error. Appendix K shows in detail how the BSV was obtained. The variance of the mean effect size $V_{\overline{d_*}}$ was estimated by taking the inverse of the summed up weights. For obtaining the standard error $SE_{\overline{d_*}}$, the variance's square root was calculated.

$$V_{\overline{d_*}} = \frac{1}{\sum_{i=1}^{k} w_i};\ SE_{\overline{d_*}} = \sqrt{V_{\overline{d_*}}} \tag{14}$$

Assuming a 95% confidence interval, the upper and lower limits ($LL_{\overline{d_*}}$ and $UL_{\overline{d_*}}$, respectively) were calculated based as follows:

$$LL_{\overline{d_*}} = \overline{d_*} - 1.96 * SE_{\overline{d_*}};\ UL_{\overline{d_*}} = \overline{d_*} + 1.96 * SE_{\overline{d_*}} \tag{15}$$

2.5.3 Binomial Effect Size Display

This section introduces a method which extends the classical interpretation option of effect sizes[13]. The so-called binomial effect size display (BESD) allows the conver-

[13] An effect size of 0.2 is interpreted as *small*, 0.5 as *medium*, and 0.8 as *high* by Cohen (1988).

sion of effect sizes into success rates which can be easily interpreted (Randolph & Edmondson, 2005; Rosenthal & Rubin, 1982). For data based on standardised mean differences, $\overline{d_*}$ has to be converted into a correlation *r* by

$$r = \frac{\overline{d_*}}{\sqrt{\overline{d_*}^2 + 4}} \tag{16}$$

Formula 16 is based on the assumption of equal sample sizes in the intervention and control group. The success rate (SR) for the two groups can then be identified as follows:

$$SR_{CG} = .5 - \frac{r}{2};\ SR_{IG} = .5 + \frac{r}{2} \tag{17}$$

For instance, a $\overline{d_*}$ of 0.30 corresponds to an *r* of .15, resulting in an increase of the success rate from 42% (control group) to 57% (intervention group). Thus, the intervention group improves 15% more, compared to the control group.

2.6 Publication Bias

Meta-analyses are at risk of resulting in non-representative findings because of selective study publication (Pigott, 2009; Rosenthal, 1979; Sterling, 1959). This problem is referred to as file-drawer problem or publication bias. It jeopardises the internal and external validity of the results by the distortion of the findings. The most prominent influential factor on publishing a scientific report is the direction or strength of its study results, with significant and large effects being more likely to be published since they are expected to be more conclusive because they are based on lager samples, compared to non-significant and small effects (Dickersin, Chan, Chalmersx, Sacks, & Smith, 1987; Dickersin, 1990; Higgins & Green, 2009; Macaskill, Walter, & Irwig, 2001). In this meta-analysis measures were taken to include published as well as unpublished reports (e.g., searching the dissertation data base ProQuest or contacting authors for requesting unpublished literature). Nevertheless, a publication bias could still influence the results. Thus, before analysing the data, methods will be applied to test the influence of a publication bias on the mean effect size.

A common method is the visual investigation by a funnel plot - a scatter plot of the study effect sizes against their sample size (Elvik, 1998). If the results are unbiased, the effect sizes should be evenly distributed around the true mean effect. If a bias exists, the effect sizes based on smaller studies would spread asymmetrically on the right side of the funnel plot. For meta-analyses which comprise of many studies, a funnel-plot might be difficult to interpret since the natural heterogeneity among the study effects already break the shape (Lipsey & Wilson, 2001). Therefore, instead of applying the traditional funnel plot, the trim and fill method by Duval and Tweedie (2000) will be utilised. It is based on the funnel plot with the difference that the Y-coordinate does show the standard error. It is also an iterative process based on symmetry assumptions: Initially, the extreme effect sizes based on small samples (corresponding to high standard errors) are being removed from the plot (trim) until the study effects are symmetrical. Then, the missing values are being integrated back into the plot (fill) together with a mirror image of each in order to correct the variance which was artificially reduced when trimming the plot. A new mean effect size based on the mirrored studies is then calculated and compared via significance test with the observed mean effect size. A publication bias is present if the two effects are significantly different. The algorithm for detecting asymmetry is sensitive to outliers, though, which is why an additional method will be introduced.

Macaskill et al. (2001) compared several methods for detecting publication bias and recommends, especially for large meta-analyses, a weighted funnel plot regression. Thus, as a second method a regression of sample size on the weighted effect size will be calculated. If the regression coefficient is significantly different from zero, the data has to be considered biased (for publication reasons).

In this meta-analysis a publication bias is rather improbable, though, since the smallest study included had to have a sample size of at least 50.

2.7 Sensitivity Analysis

When conducting a meta-analysis, many decisions need to be made which can sometimes be arbitrary, such as setting the criteria for study selection. In order to show that those decisions do not jeopardise the results of the meta-analysis, a sen-

sitivity analysis can be conducted. Hereby, the robustness of the findings are tested by re-calculating the overall effect size using alternative options where arbitrary decisions were made in the process of developing the meta-analysis. The following alternative decisions were chosen for analysing the robustness of the overall effect size in this thesis:

- **Publication Format:** No restriction was set concerning the publication format. Alternatively, only studies were included which were peer-reviewed and published.
- **Coder of the Report:** The studies were coded by three different coders. Alternatively, only scientific studies were included coded by the author of this thesis.
- **Age of Sample:** Included were scientific reports whose participants had an average age of under 18 years. Alternatively, only studies were included into the analysis if all participants are under the age of 16 years.
- **Type of Low-SES:** Concerning the low-SES criterion, studies could be included based on income, occupation, and education. Alternatively, only reports were analysed which fulfilled the low-income-SES criterion (excluding also federal assistance operationalisation).

2.8 Heterogeneity

In addition to the theoretical assumption of the random-effects model as underlying model (see section 2.5.2), the consistency between the effect sizes across the comparisons can be tested by a homogeneity test, for instance, the Q-statistics. It is based on a χ^2 distribution and tests the null hypothesis that the variability among the effect sizes is due to sampling error, thus, all comparisons evaluate the same population effect. However, the Q-test yields a p value and, thus, only offers a binary judgement. Since the aim of meta-analyses is the integration of diverse studies, substantial inconsistency should be expected so that a simple yes or no for evaluating the consistency among the results is not satisfying. Higgins, Thompson, Deeks, and Altman (2003) addressed this by introducing I^2 as a measure of the degree of heterogeneity across

studies. It "...describes the percentage of total variation across studies that is due to heterogeneity rather than chance" (Higgins et al., 2003, p. 558), calculated as follows:

$$I^2 = 100\% * \frac{Q - df}{Q} \tag{18}$$

It ranges from 0 to 100% whereas values around 25% indicate low, around 50% moderate, and around 75% high heterogeneity.

However, the index is criticised because it illustrates how much the confidence intervals overlap as opposed to really measuring the variability among the underlying effects (Borenstein et al., 2009). What needs to be kept in mind is that an I^2 of 100% does not imply a wide dispersion of the effect sizes. It allows concluding, though, that a high percentage of the observed variance is real and, thus, an investigation for moderating variables is necessary. This can be done by resorting to different procedures. The two main methods are subgroup analyses and meta-regressions.

2.8.1 Investigating Heterogeneity: Subgroup Analyses

Similar to the assumption of a model for calculating the weighted mean effects for conducting subgroup analyses (i.e., ANOVA), the type of moderator needs to be specified. They can either be *fixed* or *random*. In this context, fixed does not signify the same as in fixed-effect models. *Fixed moderators* in the context of subgroup analysis need to remain equivalent when applied in different meta-analyses. This means that if the analysis was replicated, exactly the same moderators would have to be applied. In contrast to that, random moderators can change from analysis to analysis. Concerning the investigation of moderators in this thesis, they where assumed to be fixed. The combination of presuming a random-effects model within subgroups and fixed moderators is called a mixed-effects model (MEM). For this purpose, the level of measurement has to be categorical.

2.8.2 Investigating Heterogeneity: Meta-Regression

Conducting meta-regressions is very similar to classical regressions, except for the fact that the independent and the dependent variables (as well as the possible co-

variates) are at the level of the comparisons and effect sizes, respectively, and not measured for each participants. The underlying model can be either the fixed-effect model (all comparisons have a common true effect) or the random-effects model (the effects are distributed for comparisons within the same predicted value). Here, the random-effects model will be presumed and the effect sizes estimated via methods of moments as this technique leads to the most precise results (Cooper et al., 2009; Lipsey & Wilson, 2001). In meta-regressions, both categorical as well as continuous variables can be integrated.

2.9 Applied Software

In order to scan the studies for eligibility (see section 2.2.2), the data administration programme Citavi 4 (Swiss Academic Software, 2013) was used in combination with Microsoft Excel 2007 (Microsoft, 2008). Via Citavi the study information were gathered and transferred to Excel. There, a table was compiled which contained all citations of the found studies, as well as the corresponding abstract and keywords.

For coding the publications, the handbook was transferred to the coding software FileMaker Pro 10 (FileMaker Inc., 2008). The software has many advantages compared to the classical process of filling out the coding handbook for each publication and then digitalising the data. For instance, no transcription errors can be made. For further minimising the error quote, variables can be programmed in a way that FileMaker gives feedback whenever the coder, for example, misses filling out a variable or enters an undefined number. The user interface can be individually designed and the variables clearly arranged which makes it easy to enter the data. Furthermore, by connecting the different hierarchical levels, the identification numbers are generated automatically leaving no room for mistakes. Moreover, changes are saved immediately, so that no data will be lost in case of an abnormal end. Saving changes immediately is sometimes disadvantageous if, for instance, wrong data was entered because there is no option to undo that. It is also not possible to copy-paste information across several cells which makes entering data on effect size level sometimes quite circuitous.

For analysing the data (calculating mean effectivenesses, moderator and regres-

sion analyses) the macros by D. B. Wilson (2005) were applied, together with SPSS 20 (IBM Corporation, 2011). Some analyses which could not be implemented via SPSS (e.g., trim and fill analyses and funnel plot or graphs regarding density and length, see section 3.2.1 and 3.2.5) were calculated with RStudio (The R Core Group, 2015) using the package metafor (Viechtbauer, 2010).

The dissertation was written with LaTeX using TexMaker as editor (LaTeX Project Team, 2001).

3 Results

The results of the meta-analysis are described in two main sections. The first contains descriptive results of the included scientific reports and comparisons. In the second part statistical analyses are displayed concerning the mean and long-term effectiveness of the interventions as well as the investigation of moderating variables and more differentiated analyses regarding the outcome constructs.

3.1 Descriptive Results

3.1.1 Publication Level

The 109 included studies were published in a time span from 1967 to 2013, of which the great majority of almost 85% were published within the last 15 years. Most of the reports with almost 90% were conducted in the United States. The remaining 10% were distributed among nine countries spread worldwide. Two interventions were conducted in India, Jamaica, and South Africa, respectively, and the remaining in Bahrain, Germany, Israel, Netherlands, Philippines and UK. Corresponding to the many studies conducted in the United States, almost all studies were written in English, except for one with German as publication language. Concerning the search source, screening online databases was as expected the most effective way for finding publications accounting for 70.6%. Surprisingly, findings by chance turned out to be the next successful strategy with 17.4%. This can be explained by the PrAMi-Migration online search (see section 2.2.3): 12 of the 19 studies found by chance were retrieved by this strategy. By scanning the references of the included studies 9 scientific reports could be identified. The fewest reports were discovered via scanning scientific overviews (3.7%). Considering the very wide eligibility criteria at the beginning of the project, though, which narrowed and became more specific as the project advanced, many of the initially included publications were later excluded resulting in this very small number found by the initial search strategy (3.7%).

Most of the publications (n_P = 63) were published in 42 different peer-reviewed journals. The next most common way of publishing was in form of a scientific report and a doctoral thesis with a comparable number of n_P = 20 and n_P = 17, respectively.

Few were accessible as conference proceedings (n_P = 5) or books and book chapters (n_P = 4). A detailed list of the variables can be found in Table 6.

Table 6

Descriptive Statistics of the 109 Publications

Variable	n_P	%	**Variable**	n_P	%
Year of publication			Search source		
Until 1980	8	7.3	Online search	77	70.6
1981-1990	10	9.2	Chance findings	19	17.4
1991-2000	24	22.0	Reference search	9	8.3
2001-2013	67	61.5	Scientific overviews	4	3.7
Country of performance			Organ of publication		
USA	97	89.0	Peer-reviewed journal	63	57.8
India	2	1.8	Scientific report	20	18.3
Jamaica	2	1.8	Doctoral thesis	17	15.6
South Africa	2	1.8	Conference proceeding	5	4.6
Bahrain	1	0.9	Book(chapter)	4	3.7
Germany	1	0.9	Language of publication		
Israel	1	0.9	English	108	99.1
Netherlands	1	0.9	German	1	0.9
Philippines	1	0.9			
UK	1	0.9			

Notes. n_P = number of publications.

3.1.2 Comparison Level

Of the 109 scientific reports 132 comparisons between an experimental and a control group could be coded. On the following pages central characteristics of those comparisons will be described concerning the methods applied, the samples recruited, and the interventions conducted. Tables 7, 8, and 9 show the corresponding characteristics and codings with frequencies and relative frequencies. As described

in section 1.6, one of the main research questions is to investigate the effectiveness of different types of interventions on the education-related development of children and adolescents. However, a differentiated analysis is only possible, if the number of comparisons is high enough. In this regard, the most common intervention approach was working directly with the child (child training: 61.1%, k = 82). In 28.8% of the programmes the parent was the main addressee (k = 38), or, more precisely, the mother (about 70% of parent trainings), followed by the teacher with only 5.3% (k = 7). Five interventions (3.8%) offered a combination of child and parent trainings. Thus, the great majority of the interventions (93.7%) either addressed children and/or parents. Since differentiated analyses are only feasible for child and parent trainings, all descriptive results are displayed across all comparisons and separately for child and parent trainings (subsequently referred to as CT and PT, respectively).

Methodological Characteristics. The overall minimum sample size was 50 as required by the eligibility criteria and ranged up to 12 015, with a mean of 613.70. Altogether, more than 80 000 children and adolescents across all comparisons were analysed, with a higher number in the intervention groups compared to the control groups (N_{IG} = 42 624, N_{CG} = 37 691). This is due to a control group overlap in 14 comparisons, where one control group was utilised for two or more comparisons and thus only counted once. Most of the comparisons assessed between 50 and 250 participants across IG and CG (about 66%). However, a considerate number of 20% did conduct larger-scale interventions with more than 500 participants. Random assignment to the intervention and control group was the minimum requirement for inclusion. Most authors did allocate randomly on individual level (k = 106), but some assigned whole clusters, for instance whole classes, schools, or even communities, to the intervention and control groups (about 20%). Among PTs, however, the distribution shifts by about 10% to individually allocating participants which is expectable since parents act in a naturally less clustered environment as compared to the children's school. The type of control group was quite evenly distributed among passive and active control groups, except for PTs where a higher percentage of additionally 20.7% of the comparisons included passive control groups. Hereby, active control groups contain, for example, treatment as usual or attention control groups, whereas passive control groups comprise, for instance, of untreated or wait-list con-

trol groups.)

Table 7

Descriptive Methodological Characteristics Across All Comparisons and for Child and Parent Trainings Separately

	Overall (k = 132)		Child training (k = 82)		Parent training (k = 38)	
Variable	k	%	k	%	k	%
Sample size						
50-100	48	36.4	34	41.5	13	34.2
101-250	39	29.5	25	30.5	13	34.2
251-500	17	12.9	8	9.8	6	15.8
501-1000	11	8.3	6	7.3	3	7.9
>1000	17	12.9	9	11.0	3	7.9
Randomisation						
Individual level	106	80.3	66	80.5	34	89.5
Cluster level	26	19.7	16	19.5	4	10.5
Type of control group						
Passive	63	47.7	34	41.5	26	68.4
Active	69	52.3	48	58.5	12	31.6

Notes. k = number of comparisons.

Sample Characteristics. The children's age ranged from -9 months (intervention started with pregnancy) to 19 years, with a mean age of 5 years and 11 months. The majority (about 70%) was between two and 12 years old, except for PTs with a majority of zero to 5 years of age (84.2%). The proportion of males and females was relatively evenly distributed in most of the comparison: In more than 90% the proportion of females in the samples ranges from 40 to 60% with an overall mean of 49.2%. Thus, only very few studies had a skewed gender distribution towards more females or males. Most of the comparison (about 70%) had samples that were predominated by a certain ethnicity (defined as more than 50% of the sample). The predominant

ethnicity was in about 63% of those cases either Hispanic or African-American. The high percentage of Hispanic samples is unexpected since the typical dominant ethnicity in comparable meta-analyses is African-American (e.g., Darrow, 2009; Manning et al., 2010). The PrAMi-migration search can again serve as explanation: Overall, of the 12 included comparisons which were found via PrAMi-migration search, 10 had predominant Hispanic samples. Those can not be found by a search strategy which focuses on finding poverty samples. Some comparisons consisted of primarily White participants (k_{all} = 10, k_{CT} = 4, k_{PT} = 6). There was also a substantial amount of studies (k_{all} = 21, k_{CT} = 16) with a relatively even distribution of different ethnicities. However, that was not the case for PTs with only two comparisons with a diverse sample.

Concerning the socio-economic status of the samples, the great majority of them across all comparisons and among CTs was judged as low-SES based on income (k_{all} = 109, k_{CT} = 74). However, that is not due to good income reporting (in 70.6% of all comparisons no income was reported) but can be rather explained by the availability of information concerning federal assistance measures: That is the case for 77 of the 109 comparisons. Federal assistance was primarily assessed by free- or reduced-price lunch ratings, an information easily accessible for researchers by using school data without needing to question the parents directly about their income. Some samples were judged as low-SES based on occupation and education (k_{all} = 24, k_{CT} = 6). Across all comparisons, only about every fifth or every fourth comparison, respectively, actually provided information about the occupational or educational status of the sample. This number for CTs is even lower (about 10%). In contrast to that, in evaluation reports on PTs, information on occupation and education could be obtained in about half of the comparisons and thus the inclusion based on the secondary criteria was much higher (23.7% and 21.1%, respectively). The amount of how many percent of the sample was of low-SES ranged between the defined eligibility amount of 70% and 100% with an overall mean of 88.6%. The majority of the samples almost entirely consisted of participants from low-SES (category *91 to 100%*: k_{all} = 68, k_{CT} = 45, k_{PT} = 19).

Additionally to being of low-SES, further risk factors could have occurred. Thereby, it was distinguished between child-related risk factors (such as academic problems,

conduct problems, or low birth weight) and parental risks factors (such as mother's low IQ, abusiveness toward child, or being divorced). The risk factor applied if at least 50% of the sample was affected. About 30% of the samples were characterised by either child-related or parental-related additional risk with a distribution shift for CTs towards child-related risks (29.3%) and for PTs towards parent-related risks (21.1%). Participants of only very few comparisons were at risk concerning both criteria (range between 0 and 5.3%).

In a nutshell, the typical sample in this meta-analysis was between two and 12 years of age with an even gender distribution, consisting predominantly of African-Americans or Hispanics who almost all were of low-SES not affected by additional risk factors. Among PTs, the typical sample was between 0 and 5 years of age.

Table 8

Descriptive Sample Characteristics Across All Comparisons and for Child and Parent Trainings Separately

	Overall (k = 132)		Child training (k = 82)		Parent training (k = 38)	
Variable	k	%	k	%	k	%
Age distribution (in months)						
<0	5	3.8	1	1.2	4	10.5
0-24	20	15.2	3	3.7	16	42.1
25-60	39	29.5	24	29.3	12	31.6
60-144	53	40.2	43	52.4	5	13.2
>144	15	11.4	11	13.4	1	2.6
Gender distribution						
Predominantly male	5	3.8	2	2.4	2	5.3
Evenly distributed	125	94.7	78	95.1	36	94.7
Predominantly female	2	1.5	2	2.4	–	–
Dominant ethnicity						
Hispanic	42	31.8	25	30.5	11	28.9
African-American	41	31.1	28	34.1	11	28.9

continued on the next page...

Table 8

Descriptive Sample Characteristics Across All Comparisons and for Child and Parent Trainings Separately (continued)

	Overall (k = 132)		Child training (k = 82)		Parent training (k = 38)	
Variable	k	%	k	%	k	%
Diverse	21	15.9	16	19.5	2	5.3
White	10	7.6	4	4.9	6	15.8
Not reported	18	13.6	9	11.0	8	21.1
Type of low-SES[a]						
Income	109	82.6	74	90.2	23	60.5
Occupation	16	12.1	6	7.3	9	23.7
Education	8	6.1	–	–	8	21.1
Other	10	7.6	3	3.7	7	18.4
Low-SES distribution (in %)						
70-80	39	29.5	22	26.8	13	34.2
81-90	25	18.9	15	18.3	6	15.8
91-100	68	51.5	45	54.9	19	50.0
Risk factors						
No	90	68.2	53	64.6	27	71.1
Child	26	19.7	24	29.3	1	2.6
Parental	14	10.6	5	6.1	8	21.1
Both	2	1.5	–	–	2	5.3

Notes. k = number of comparisons.
[a] Variable does not sum up to k_{all} = 132, k_{CT} = 82, and/or k_{PT} = 38 because several criteria could apply per comparison.

Intervention Characteristics. Descriptive characteristics concerning the intervention are displayed in Table 9. The variable *content of intervention* was divided in three sections by addressee (a. child, b. parent, or c. other party was approached) which was further subdivided by topic of the particular training (multiple responses across but not within the three sections were allowed). The majority of the comparisons (68.9%) included some form of training the children's (meta-)linguistic skills and/or cognitive/learning skills (all of the 33 comparisons with more than one approach un-

der a) conducted at least one of the just mentioned). The second greatest group practised constructive parenting accounting for more than one third of the comparisons (36.4%, among the category *More than one approach* all included parenting), followed by approaches addressing the school as context (15.2%). Certainly, among child and parent trainings all comparison trained child or parental skills, respectively. In some comparisons categorised as CTs (k_{all} = 13), the teaching staff was also included. As for PTs half of the interventions included also child contents.

Over all comparisons, the most often applied method with 59.8% was doing exercises like writing essays, doing maths calculations, or giving homework. Performance and the disposal of enriching materials was used in 35.6% and 32.6%, respectively. Methods applied in every fifth to fourth intervention were group discussions, home visits, psycho-education, information transfer, and behavioural training with instructions. In almost 77% of the CTs exercises were conducted, followed by performances (37.9%) and enriching materials (25.6%). In about every fifth CT each of the following methods was applied: group discussions, information transfer, and free/guided play. As for PTs, most of the comparisons included home visitation (65.8%) or/and behavioural training (50.0%). More than one third of the PTs resorted to one or more of the following methods: exercises, performances, enriching materials, groups discussions, psycho-education, or information transfer. The majority of the programmes were conducted with four or less different methods (about 80%). However, among PTs most programmes used three or four methods, whereas among CTs mostly one or tow different ones are applied. This could be observed already in the variable *methods of intervention* where across all child trainings only four methods are conducted in more than 20% of the comparisons as opposed to eight among PTs. Across all comparisons, about half of the trainings were delivered in a group setting, whereas 39 were conducted on an individual basis and 24 were realised in both formats. Here, the distribution among CTs and PTs differed considerably compared to all comparisons: CTs were primarily realised in a group setting (72.0%) whereas PTs were most often conducted on an individual base (60.5%) or in both formats (23.7%). Typically a manual or curriculum was applied (about 76%) which corresponds to the high percentage of structured programmes (between 63.2 and 84.1%). PTs, however, were in 36.8% of the cases of low structure. Most of the interventions were adaptable to the individual's

needs or characteristics (between 60.5 and 67.1%) with a very high level of participatory programme parts (between 87.9 and 92.1%). Regarding all comparisons, the duration of the programmes was diverse. While child trainings were mostly realised within three to 12 months (51.2%), PTs most often took place for more than 12 months (47.4%). Less than one third of the programmes were culturally tailored, except for PTs where the number was higher (42.1%). The trainer's professional background was diverse, only CTs were conducted primarily by educational staff (48.8%).

Across all comparisons, the intervention setting was primarily the educational facility of the child (k_{all} = 82), followed by the home of the parents with 22 comparisons. Not surprisingly, child trainings were primarily realised in an educational facility (86.6%) and PTs at home (52.6%). The majority of the interventions was conducted in an urban environment (between 52.6 and 59.8%) which was not described as particularly socially deprived (i.e., ghetto, social hotspot, low-income area or the like; between 60.5 and 69.5%).

Summarised, the typical intervention across all comparisons was training the language/cognitive/learning skills of the child by means of exercises, performances, and/or enriching materials in a group setting with a structured curriculum, however, adaptable to an individual's needs with many participatory possibilities except for cultural tailoring. The length of the programmes as well as the professional background of the trainers were rather diverse. The interventions usually took place in an educational facility within an urban and non-deprived environment. In contrast, among CTs usually only one or two different methods were applied, typically exercises, whereas PTs were realised by implementing three or four, typically home visits and/or behavioural training. Moreover, PTs typically focused on the topic parenting, were more structured, more often culturally tailored and conducted over a longer period of time, in comparison to CTs. Child trainings were, in addition, primarily implemented by educational staff.

Table 9

Descriptive Intervention Characteristics Across All Comparisons and for Child and Parent Trainings Separately

	Overall (k = 132)		Child training (k = 82)		Parent training (k = 38)	
Variable	k	%	k	%	k	%
Content of intervention[a]						
a) Child development approach						
Language	39	29.5	33	40.2	6	15.8
Cognition/learning	19	14.4	15	18.3	4	10.5
Psychosocial	10	7.6	8	9.8	–	–
General development	4	3.0	2	2.4	1	2.6
More than one approach	33	25.0	24	29.3	8	21.1
b) Parent skill approach						
Parenting	33	25.0	5	6.1	26	68.4
Social support	3	2.3	–	–	3	7.9
More than one approach	15	11.4	3	3.7	9	23.7
c) Approaching other parties/contexts						
Teacher/school	20	15.2	13	15.9	–	–
Neighbourhood	1	0.8	–	–	1	2.6
More than one approach	1	0.8	1	1.2	–	–
Methods of intervention[a]						
Exercises	79	59.8	63	76.8	14	36.8
Performance	47	35.6	31	37.9	14	36.8
Enriching material	43	32.6	21	25.6	16	42.1
Group discussion	33	25.0	17	20.7	13	34.2
Home visitation	31	23.5	3	3.7	25	65.8
Psycho-education	31	23.5	8	9.8	15	39.5
Information transfer	31	23.5	16	19.5	13	34.2
Behavioural training	31	23.5	8	9.8	19	50.0
Free or guided play	19	14.4	15	18.3	3	7.9
Joint activities	14	10.6	7	8.5	4	10.5
Process diagnostic	10	7.6	9	11.0	1	2.6
Networking	9	6.8	4	4.9	3	7.9

continued on the next page...

Table 9

Descriptive Intervention Characteristics Across All Comparisons and for Child and Parent Trainings Separately (continued)

	Overall (k = 132)		Child training (k = 82)		Parent training (k = 38)	
Variable	k	%	k	%	k	%
Other	8	6.1	5	6.1	2	5.3
Number of methods[b]						
One to two	56	42.4	45	54.9	6	15.8
Three to four	56	42.4	29	35.4	24	63.2
Five and more	19	14.4	8	9.8	8	21.1
Delivery format[b]						
Individual	39	29.5	15	18.3	23	60.5
Group	68	51.5	59	72.0	6	15.8
Group & individual	24	18.2	8	9.8	9	23.7
Application of a manual[b]						
No	30	22.7	17	20.7	11	28.9
Manual/curriculum	100	75.8	65	79.3	27	71.1
Structuredness[b]						
Low	29	22.0	13	15.9	14	36.8
High	101	76.5	69	84.1	24	63.2
Individualisation[b]						
Low	48	36.4	27	32.9	15	39.5
High	82	62.1	55	67.1	23	60.5
Level of participation[b]						
Low	15	11.4	9	11.0	3	7.9
High	116	87.9	73	89.0	35	92.1
Length of programme						
Up to three months	41	31.1	29	35.4	9	23.7
Three to 12 months	59	44.7	42	51.2	11	28.9
More than 12 months	32	24.2	11	13.4	18	47.4
Cultural tailoring[b]						
No	92	69.7	64	78.0	22	57.9
Yes	38	28.8	18	22.0	16	42.1

continued on the next page...

Table 9

Descriptive Intervention Characteristics Across All Comparisons and for Child and Parent Trainings Separately (continued)

	Overall (k = 132)		Child training (k = 82)		Parent training (k = 38)	
Variable	k	%	k	%	k	%
Profession of the trainer						
Educational staff	46	34.8	40	48.8	5	13.2
University staff	23	17.4	14	17.1	9	23.7
Psychosocial professional	22	16.7	6	7.3	11	28.9
Multiple trainers	17	12.9	8	9.8	5	13.2
Layperson	14	10.6	7	8.5	7	18.4
Not in person	5	3.8	5	6.1	–	–
Clinical professional	1	0.8	1	1.2	–	–
Not specified	4	3.0	1	1.2	1	2.6
Environment of intervention						
Setting of intervention[b]						
Educational facility	82	62.1	71	86.6	4	10.5
Home	22	16.7	2	2.4	20	52.6
Other	14	10.6	6	7.3	6	15.8
Several settings	12	9.1	3	3.7	7	18.4
Area of intervention						
Rural/suburban	23	17.4	16	19.5	6	15.8
Urban	77	58.3	49	59.8	20	52.6
Mixed	22	16.7	11	13.4	8	21.1
Not specified	10	7.6	6	7.3	4	10.5
Socially deprived area						
No	90	68.2	57	69.5	23	60.5
Yes	42	31.8	25	30.5	15	39.5

Notes. k = number of comparisons.

[a] Variable does not sum up to k_{all} = 132, k_{CT} = 82, and/or k_{PT} = 38 because several criteria could apply per comparison.

[b] Variable does not sum up to k_{all} = 132, k_{CT} = 82, or k_{PT} = 38 because of one or two missings.

3.2 Statistical Analyses

The statistical analyses will be conducted in six parts in accordance with the structure of the research question (see section 1.6). First, the data will be investigated for potential publication bias (see section 3.2.1), followed by displaying mean effectiveness and the long-term stability of the effects for the main outcome categories, as well as the short-term effectiveness by intervention type (see sections 3.2.2, 3.2.3, and 3.2.4). Then, the variance among the study effect sizes will be analysed via moderator analyses (see section 3.2.5). In a last step, a differentiated outcome analyses with relevant moderators will be conducted regarding the two child outcome constructs basic development and school development (see section 3.2.6).

In the previous chapter the overall mean difference weighted by the inverse variance was indicated with $\overline{d_*}$. In order to simplify this description in will be referred to as d in this chapter.

3.2.1 Publication Bias

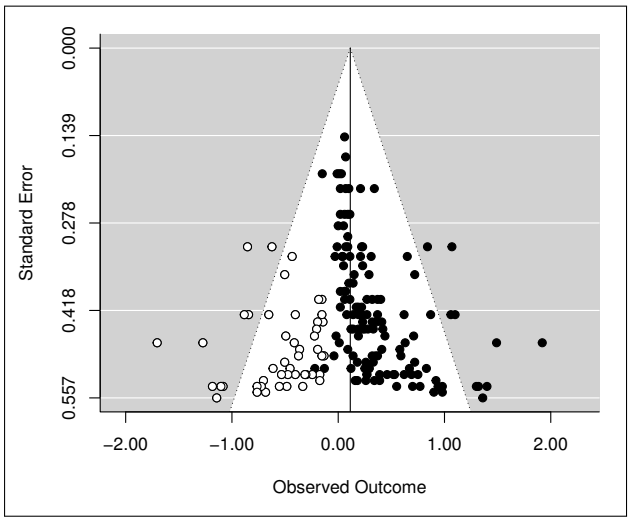
As mentioned in section 2.6 the influence of a publication bias on the mean effect size has to be investigated before continuing with further analyses. As depicted in Figure 4, the observed effect sizes (displayed as black circles) are distributed strongly asymmetric to the right side of the funnel plot leaving a gap of small studies with negative effects. This seems like a text book example of publication bias. The first visual evaluation is supported by the trim and fill analyses where the adjusted mean effect (d_{adj} = 0.11) is significantly smaller than the original one (d_{obs} = 0.31) as can be seen by the non-overlapping confidence intervals. Forty-seven effect sizes had to be added (white circles) to establish symmetry. Additionally, the regression slope of the weighted funnel plot regression differs significantly from zero (β = -0.28, $p < .001$), indicating a decrease in effect size with increasing sample sizes.

This indicates that the adjusted mean effect size should be interpreted, rather than the observed one. However, it is a surprising finding, particularly since a publication bias was not expected due to the minimum sample size of 50 which a study needed to have in order to be included. Since all three approaches are based on the assumption that large studies shouldn't be affected by selective publication processes, the

three methods where applied only on studies with a sample size of more than 100 (k = 84). The number 100 is rather arbitrary but high enough not to be considered as small. As can be seen in Figure 5, the picture does not change: A skewed distribution and significantly different original versus mirrored effect sizes, confirmed by a significant regression slope (β = -0.26, p < .001). These results imply other causes than publication bias for the asymmetric distribution of the effect sizes when predicted by sample size. Consequently, the results will be regarded as unbiased and sample size will be further investigated in section 3.2.5 (paragraph ***Methodological Moderators***) together with other moderating variables. It is, for instance, presumed that sample size could function as a proxy for implementation quality (Beelmann, 2006; Egger, Smith, Schneider, & Minder, 1997).

Figure 4

Trim and Fill Analysis: Funnel Plot Across all Comparisons

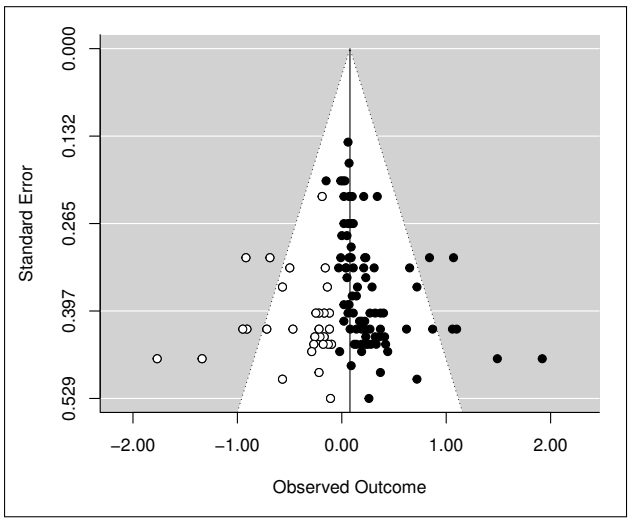


	d [95% CI]	k
Observed	0.31^{***} [0.26, 0.35]	132
Adjusted	0.11^{***} [0.05, 0.18]	179

Notes. On the X-axis the effect sizes on comparison level are displayed and the standard error on the Y-axis. The observed effect sizes are depicted as black circles, whereas the imputed ones are depicted as white circles. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Figure 5

Trim and Fill Analysis: Funnel Plot for Large Comparisons (n > 100)



	d [95% CI]	k
Observed	0.24*** [0.19, 0.29]	84
Adjusted	0.08* [0.01, 0.15]	114

Notes. Large On the X-axis the effect sizes on comparison level are displayed and the standard error on the Y-axis. The observed effect sizes are depicted as black circles, whereas the imputed ones are depicted as white circles.

*$p < .05$, **$p < .01$, ***$p < .001$.

3.2.2 Mean Effectiveness and its Robustness

The observed weighted mean effect size across all comparisons is $d = 0.31$ ($SE = 0.03$, $p < .001$) with the lowest effect size of -0.22 and the highest of 1.92. According to Cohen (1988), the overall effect can be interpreted as small and corresponds to a 15.3% increase of the success rate from the intervention to the control group. The significant Q-statistic supports the theoretical assumption (see section 2.5.2) of a random-effects model ($Q_B = 737.83$, $df = 131$, $p < .001$). The high proportion of true variance ($I^2 = 82.2\%$) suggests the existence of one or several variable(s) which moderate the mean effect. If the results were biased due to selective publication processes, the trim and fill-adjusted effect size of $d = 0.11$ ($SE = 0.03$, $p < .001$) would have been smaller but still significantly different from zero.

Sensitivity Analysis. In order to test the robustness of the mean effect size, sensitivity analyses were calculated. As can be retrieved from Table 10 by comparing the confidence intervals of the original effect size with each of the ones retrieved based on alternative criteria, they all overlap. Thus, none of the newly derived effects were

significantly different from the original average effect size.

Table 10

Sensitivity Analyses: Mean Effect Sizes when Changing Decisions

Alternative Criterion	d	[95% CI]	k
Publication format: Peer-reviewed and published	0.29***	[0.24, 0.34]	114
Coder of the Report: Only thesis author	0.31***	[0.25, 0.36]	99
Maximum Age of Sample: Below 16 years	0.31***	[0.26, 0.36]	120
Type of Low-SES: Low-income-SES	0.31***	[0.19, 0.44]	32
Original mean effect size	0.31***	[0.26, 0.35]	132

Notes. d = average mean difference weighted by the inverse variance, CI = confidence interval, k = number of comparisons.
*$p < .05$, **$p < .01$, ***$p < .001$.

3.2.3 Long-Term Effectiveness

Table 11 shows the mean effectiveness over time and separately for the three outcome categories. The outcome categories will subsequently be abbreviated as follows: CO for child outcomes, PO for parental outcomes, and SO for school outcomes.

The mean effect over all categories (first column on the right in Table 11) was stable on a small level (d_{short} = 0.32, $p < .001$; d_{mid} = 0.36, $p < .001$) until 12 months after the intervention ended when it decreased significantly. It was still significantly different from zero but only reached a very low value (d_{long} = 0.12, $p = .003$).

When looking at the short-term effectiveness of the interventions, the mean effect sizes for each of the three outcome categories can be interpreted as small but significantly different from zero (d_{CO} = 0.31, $p < .001$; d_{PO} = 0.27, $p < .001$; d_{SO} = 0.27, $p < .001$). This changes for POs if they were measured more than 12 months after the intervention ended: The intervention then has no effect any more (d_{long} = -0.03, $p = .386$). For SOs the intervention effect already vanishes after three months post intervention (d_{mid} = -0.01, $p = .906$). The most lasting effect can be found for COs: Even though the effect diminishes significantly from the mid- to the long-term period, it is still substantial (d_{mid} = 0.35, $p < .001$; d_{long} = 0.11, $p = .002$).

Since short-term effectiveness could be found for all three categories based on a large enough number of comparisons, further analyses will focus on short-term outcomes only. Furthermore, it can be concluded from Table 11, that the effectiveness for each category across all time intervals declines descriptively from child ($d = 0.29$, $p < .001$) to parental ($d = 0.22$, $p < .001$) to school outcomes ($d = 0.17$, $p < .001$). This difference is not statistically significant, however, it raises the question of a differentiated analysis by main addressee of the intervention (main addressee can be either the child, the parents, the teacher, or a combination of children and parents).

Table 11

Mean Effectiveness by Time-Point and Outcome Category

	Outcomes											
	Child			Parental			School			Overall		
Time interval	d [95% CI]	k	I^2	d [95% CI]	k	I^2	d [95% CI]	k	I^2	d [95% CI]	k	I^2
Short-term	0.31***° [0.26, 0.36]	132	84.3%	0.27***° [0.18, 0.37]	30	65.6%	0.27***° [0.13, 0.40]	10	53.1%	0.32***° [0.27, 0.37]	132	83.6%
Mid-term	0.35***° [0.21, 0.48]	24	87.1%	0.22*° [0.01, 0.44]	8	81.2%	−0.01 [−0.10, 0.08]	3	0%	0.36***° [0.22, 0.50]	24	84.8%
Long-term	0.11**° [0.04, 0.18]	24	42.9%	−0.03 [−0.09, 0.04]	8	0%	−0.02 [−0.11, 0.07]	3	0%	0.12**° [0.04, 0.19]	24	47.0%
Overall	0.29***° [0.25, 0.34]	132	82.6%	0.22***° [0.13, 0.31]	34	63.3%	0.17**° [0.05, 0.28]	13	61.8%	0.31***° [0.26, 0.35]	132	82.2%

Notes. d = average mean difference weighted by the inverse variance, CI = confidence interval, k = number of comparisons, I^2 = heterogeneity.
*$p < .05$, **$p < .01$, ***$p < .001$, °significant heterogeneity at the 5% significance level.

3.2.4 Short-Term Effectiveness by Intervention Type

In Table 12 (first column on the right), the short-term effects by type of intervention and outcome category are displayed. CTs reach higher descriptive overall effects (d = 0.35, p < .001), followed by PTs (d = 0.27, p < .001), followed by teacher trainings (TTs) (d = 0.18, p < .001). However, the differences only obtain significance for child compared to teacher trainings as the non-overlapping confidence intervals indicate. Child and parent trainings combined have the highest average and child outcome effect size, both almost reaching medium height with d_{all} = 0.45 and with d_{CO} = 0.44, however, it only marginally reaches statistical significance (p = .085). This can be explained by the high range of the effects from -0.01 to 1.24. Furthermore, the mean effect sizes tend to be largest within one training type if the type of training and outcome category correspond: CTs have the highest average effect sizes on CO with d = 0.35 (p < .001, abbreviated by CT/CO hereafter), parent trainings on parent outcomes with d = 0.35 (p < .001, abbreviated by PT/PO hereafter) and teacher trainings on school outcomes with d = 0.25 (p < .001, abbreviated by TT/SO hereafter). These effects are all small and significant. Important to point out is that the CT/CO average effect size is descriptively not the highest among CTs, instead CT/SO is with d = 0.48. This effect consists of highly varying effect sizes (I^2 = 75.6%), though, is not statistically different from zero (p = .087), and additionally based on only three comparisons. What further surprises is the comparably high effectiveness of PTs on COs with d = 0.27 (p < .001, abbreviated by PT/CO hereafter) as compared to CT/COs (d = 0.35, p < .001). Since those two training types will be contrasted in detail in further analyses regarding child outcome effectiveness, a significance test between CT/COs versus PT/COs was conducted, yielding a non-significant result (Q_B = 1.77, df = 1, p = .183). No other significant effects could be found or the effect sizes were based on less than five comparison and are, thus, not credible.

Heterogeneity. Concerning the heterogeneity, the significant Q-statistics for CT/CO, PT/CO, and PT/PO combinations indicate the existence of true variance among the effect sizes expressed in moderate to high I^2 values which will be investigated by moderator analyses. As far as variability among TT/SO effects is concerned, the non-significant Q-statistic indicates that the true dispersion among the effects is zero. In contrast to that, I^2 is of moderate height. Nevertheless, no further investiga-

tions for moderating variables can be conducted since the number of comparisons is too low. That is also the case for combined programmes.

Long-term Effectiveness for CTs and PTs. Since the meta-analysis focuses on investigating the effectiveness on COs by contrasting child and parent trainings, the stability of the effects for those two programme types was analysed in detail. Regarding CTs, the small short-term effect (d_{short} = 0.35, $p < .001$) descriptively increases to medium size when child outcomes were measured between three to 12 months after the intervention ended (d_{mid} = 0.57, $p < .001$), but levels back to small size regarding the long-term time interval (d_{long} = 0.24, $p = .001$). The PT effect on child outcomes stays relatively stable on a small level regarding short- (d_{short} = 0.27, $p < .001$) and mid-term time interval (d_{mid} = 0.26, $p = .008$) and drops to zero when analysing long-term outcomes only (d_{long} = 0.08, $p = .139$).

Table 12

Short-Term Mean Effectiveness Intervention and Outcome Category

	Outcomes											
	Child			Parental			School			Overall		
Training	*d* [95% CI]	*k*	I^2	*d* [95% CI]	*k*	I^2	*d* [95% CI]	*k*	I^2	*d* [95% CI]	*k*	I^2
Child	0.35***° [0.29, 0.41]	82	83.0%	0.07 [−0.09, 0.23]	3	48.7%	0.48° [−0.07, 1.04]	3	75.6%	0.35***° [0.29, 0.42]	82	83.4%
Parent	0.27***° [0.18, 0.36]	38	76.9%	0.35***° [0.24, 0.46]	24	47.6%	–	–	–	0.27***° [0.18, 0.35]	38	74.6%
Teacher	0.05° [−0.02, 0.12]	7	64.3%	0.06 [−0.07, 0.20]	1	–	0.25*** [0.10, 0.40]	7	42.9%	0.18***° [0.07, 0.28]	7	61.3%
Combination	0.44° [−0.08, 0.96]	5	95.3%	0.04 [−0.15, 0.23]	2	0%	–	–	–	0.45° [−0.06, 0.96]	5	95.3%

Notes. d = average mean difference weighted by the inverse variance, CI = confidence interval, k = number of comparisons, I^2 = heterogeneity.
*$p < .05$, **$p < .01$, ***$p < .001$, °significant heterogeneity at the 5% significance level.

Table 13

Long-Term Mean Effectiveness of Child and Parent Trainings for Child Outcomes

	Child Training			Parent Training		
	d			d		
Time interval	[95% CI]	k	I^2	[95% CI]	k	I^2
Short-term	0.35***° [0.28, 0.41]	82	83.0%	0.27***° [0.18, 0.36]	38	76.9%
Mid-term	0.57***° [0.26, 0.89]	11	91.4%	0.26**° [0.07, 0.46]	8	60.4%
Long-term	0.24**° [0.10, 0.38]	12	63.7%	0.08 [−0.02, 0.18]	8	0%

Notes. d = average mean difference weighted by the inverse variance, CI = confidence interval, k = number of comparisons, I^2 = heterogeneity.
*$p < .05$, **$p < .01$, ***$p < .001$, °significant heterogeneity at the 5% significance level.

3.2.5 Confoundation of the Mean Effect

In order to identify relevant moderating variables, subgroup analyses (ANOVAs) and meta-regressions were conducted. The moderating influence was calculated concerning overall outcomes as well as across child outcomes (CO) for child trainings (CTs), and across child and parent outcomes for parent trainings (PTs). Only effect sizes which were assessed within three months after the intervention ended were included. Correspondingly to the structure of the descriptive results on the comparison level (see section 3.1.2), the moderators were grouped in three types and the results are depicted accordingly: methodological moderators (see Table 14), sample moderators (see Table 16), and intervention moderators (see Table 17). Detailed information regarding the analysed moderators can be found in Appendix H. All analyses were based on the assumption of a

Continuous variables were analysed both via meta-regressions and ANOVAs (as categorised variables). In doing so, clarity was ensured and no information due to categorisation was lost. The corresponding meta-regression was placed directly below

the subgroups analysis in Tables 14, 16, and 17, showing the standardised regression coefficient β and the explained variance R^2 of the model. Since exclusively simple linear regressions are depicted in those tables, the significance test of the model complies with the one of the regression coefficient. Hence, only the regression coefficient and its significance will be reported.

The analysing strategy for Table 14, 16, and 17 will be as follows: First, the effect of PTs on POs will be summarised briefly and then a deeper look will be taken when analysing the moderation effect on COs by contrasting child and parent trainings. Then the non-moderating variables will be mentioned. Thus, the focus will be set primarily on comparing how child outcome effectiveness was moderated among CTs versus PTs. Finally, a summary of the paragraph will be given. Concerning methodological moderator analyses, though, publication bias will be analysed before continuing with the just introduced structure.

Methodological Moderators

Publication Bias. As stated already in section 3.2.1, the moderating effect of sample size needs to be investigated in detail. In Table 14, an interesting pattern can be observed: The predictive value of sample size for the average effect size exists among COs (CT/CO: β = -0.26, p = .003; PT/CO: β = -0.44, p = .001) whereas sample size shows to have no moderating effect concerning POs (β = 0.08, p = .700). Thus, POs seem not to be affected by publication bias. Since it is not reasonable that a publication bias exists for child but not for parental outcomes, this can be taken as additional hint for an explanation other than publication bias for the association of sample size and effect size.

In order to further enquire this possible publication bias it would be worth concerning any variable associated with sample size as well as effect size. Recommendations by Egger et al. (1997) who describe several possible sources of asymmetry in funnel plots where taken into account, for instance, methodological quality (smaller studies tend to be less rigorous), risk characteristics (smaller studies are more likely to be conducted with children at risk), and intervention characteristics (smaller studies are better implemented and more intense). Additionally, the area where the programme was conducted was expected to be associated with sample size with larger programmes being conducted in mixed settings (urban and rural) as opposed to exclusively urban

or rural areas. Furthermore, it was considered whether the intervention was conducted as a single- or a multi-site study expecting larger programmes to be realised in multiple sites[14].

The selected variables were included each as additional regressor to sample size for predicting child effect sizes, separately for CTs and PTs in order to consider the possibility of different explanations for each type of intervention. It was expected that sample size should loose its significant association with effect size to the added predictor. As can be seen in Table 15, the predictive value of sample size was exceeded by *area of conduct* among CTs (β_n = -0.12, p = 0.27; β_{area} = -0.26, p = .018) and *site of study* among PTs (β_n = -0.15, p = .360; β_{site} = -0.46, p = .007). Thus, it can be concluded that the association between sample size and effect size is not due to publication bias. Furthermore, among CTs area of conduct was associated with *supervision of the trainer* ($\chi^2[1, k = 72] = 4.58$, p = .034) with only half as many supervised intervention having been conducted in rural or urban areas (43.5%) compared to mixed areas (80.0%). *Site of study* covaried with *manualisation* ($\chi^2[1, k = 38] = 5.55$, p = .031) as well as with *deprivation of the area* ($\chi^2[1, k = 38] = 6.61$, p = .010): Among single-site comparisons, 80.0% did apply a curriculum or manual in contrast to only 37.5% among multi-site comparisons and 50% of the comparisons within single-site studies were implemented in socially deprived areas (e.g., low-SES neighbourhood) as opposed to none within multi-site ones.

Next, the remaining methodological variables were investigated regarding moderating effects. The presented results are displayed in Table 14.

Parent Trainings on Parental Outcomes (PT/PO). Among PTs on POs no significant moderating influence could be found regarding methodological moderators.

Child and Parent Trainings on Child Outcomes (CT/CO versus PT/CO). Regarding COs, a significant moderating effect of *group equivalence* among CTs but not among PTs could be observed (CT: Q_B = 7.64, df = 1, p = .006; PT: Q_B = 1.14, df = 1, p = .286). Further, a non-significant moderation among CTs but a significant one for PTs of the variable *contact between IG/CG* could be observed (CT: Q_B = 0.01, df = 1,

[14] A single-site study takes place usually in one place (e.g. city or county), whereas multi-site studies are realised across a wide geographical area (e.g. several states of a country).

p = .979; PT: Q_B = 3.91, df = 1, p = .048).

As for the moderator group equivalence among CTs, only those comparisons were included which tested pre-test differences regarding demographic (e.g., pre-test age) or assessment relevant variables (e.g., pre-test IQ score). If the groups differed at pretest the average post effect appeared to be larger with d = 0.54 (p < .001) as opposed to a smaller effect of d = 0.26 (p < .001) within equivalent groups. Group equivalence was found to be associated with *type of randomisation* (χ^2[1, k = 49] = 11.75, p = .002): Higher percentages of individually randomised samples accumulated among equivalent groups (86.8%) compared to non-equivalent groups (36.4%). The moderating effect of group equivalence disappeared if analysed separately for individual- versus cluster-randomised comparisons (individual: Q_B = 0.20, df = 1, p = .657; cluster: Q_B = 3.21, df = 1, p = .073). As mentioned above, for the combination PT/CO this moderator had no effect and neither was associated with type of randomisation (χ^2[1, k = 28] = 3.29, p = .135).

Regarding contact between the two groups during intervention conduct, the direction of the effect among PTs on COs was surprising. If intervention and control group participants did interact the average effect size was higher (d = 0.39, p < .001) than if they had no contact (d = 0.20, p < .001). One explanation could be the confoundation of contact with social deprivation of the area (χ^2[1, k = 38] = 14.18, p < .001). If the groups interacted it was more likely that the intervention took place within a socially deprived region (78.6%) than if there was no contact (16.7%). If the intervention was realised within socially deprivation areas, in turn, higher effectiveness could be observed (see Table 17, Q_B = 8.70, df = 1, p = .003). By including both moderators into a meta-regression on child outcomes, the remaining predictive value considering the covariation could be analysed. The results revealed that social deprivation of the area was marginally significant (β = 0.35, p = .056), whereas contact had no predictive effect anymore (β = 0.09, p = .619).

No moderating effect could be found for *type of randomisation*, *type of control group*,

analysis approach[15], *attrition rate*[16], and *attendance rate*[17] as can be seen in Table 14.

In summary, four important results should be noted here:

1. **Publication Bias:** The association of sample size and effect size disappeared if area of conduct among child trainings and site of study among parent trainings were included in the regression together with sample size. Studies conducted in mixed areas were about twice as likely to supervise the trainer compared to just rural or urban areas. It was also more likely that single-site studies were manualised and implemented within socially deprived areas as opposed to multi-site studies.

2. **PT/PO:** No methodological variable significantly moderated the effectiveness among parent trainings on parental outcomes.

3. **CT/CO and PT/CO - Group Equivalence:** Among child trainings it could be observed that comparisons with pretest differences between the intervention and control group reached a significant medium effect size ($d = 0.54$) whereas equivalent ones were lower but still significantly different from zero ($d = 0.26$). Among equivalent groups more individually randomised trials could be found (about 87%) than among non-equivalent ones (about 37%). The moderating effect of group equivalence vanished if tested separately for individual- and cluster-randomised studies. Group equivalence did not significantly influence parent training effectiveness regarding child outcomes.

4. **CT/CO and PT/CO - Contact IG/CG:** Parent training effectiveness on child outcomes appeared to be higher if the intervention and control group interacted ($d = 0.39$) as if they didn't ($d = 0.20$). It should be mentioned that comparisons

[15] Analysis approach contrasts the inclusion of only analysing participants who received a certain amount of the program into calculating the intervention effect (selected) versus the inclusion of all participants who were randomised or who attended the post-test (universal).

[16] The ratio of participants who did not attend the post-test to those who completed randomisation procedure.

[17] The ratio of the mean attended number of sessions to the maximal number of session.

with interacting intervention and control groups were more likely to be conducted in socially deprived areas than comparisons with no contact. Further, if both variables were included into one meta-regression on child outcomes the predictive effect of contact disappeared. Regarding child trainings, no significant difference was found whether the groups interacted or not.

Table 14

Methodological Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Sample size			Q_B = 29.92***			Q_B = 0.18			Q_B = 13.20**			Q_B = 32.18***
50 to 250	0.46***	59	[0.39, 0.54]	0.33***	17	[0.19, 0.47]	0.34***	26	[0.24, 0.44]	0.42***	87	[0.35, 0.48]
251 to 1000	0.24***	14	[0.12, 0.35]	0.38***	7	[0.20, 0.57]	0.24***	9	[0.11, 0.37]	0.29***	28	[0.21, 0.38]
More than 1000	0.08	9	[−0.047, 0.21]	–	–	–	−0.04	3	[−0.21, 0.14]	0.09	17	[−0.01, 0.18]
Meta-regression	−0.26**	82	.07	0.08	24	.01	−0.44**	38	.20	−0.28***	132	.08
Type of randomisation			Q_B = 1.73			–			Q_B = 3.78			Q_B = 0.03
Individual level	0.33***	66	[0.25, 0.41]	–	–	–	0.29***	34	[0.20, 0.38]	0.33***	106	[0.27, 0.38]
Cluster level	0.44***	16	[0.30, 0.58]	–	–	–	0.05	4	[−0.18, 0.28]	0.32***	26	[0.21, 0.42]
Type of CG			Q_B = 0.60			Q_B = 1.61			Q_B = 0.27			Q_B = 0.06
Passive	0.38***	34	[0.28, 0.48]	0.31***	18	[0.19, 0.44]	0.26***	26	[0.14, 0.37]	0.33***	63	[0.26, 0.40]
Active	0.33***	48	[0.24, 0.42]	0.48***	6	[0.25, 0.71]	0.31***	12	[0.14, 0.48]	0.32***	69	[0.25, 0.39]
Contact IG/CG			Q_B = 0.01			Q_B = 3.52			Q_B = 3.91*			Q_B = 7.91**
No	0.35***	10	[0.17, 0.54]	0.26***	14	[0.12, 0.40]	0.20***	24	[0.09, 0.31]	0.22***	42	[0.14, 0.31]
Yes	0.35***	72	[0.28, 0.42]	0.46***	10	[0.30, 0.62]	0.39***	14	[0.24, 0.54]	0.38***	90	[0.31, 0.44]
Group equivalence			Q_B = 7.64**			Q_B = 0.67			Q_B = 1.14			Q_B = 1.94
Not given	0.54***	11	[0.37, 0.72]	0.68	1	[−0.04, 1.41]	0.10	5	[−0.09, 0.30]	0.40***	17	[0.25, 0.55]
Given	0.26***	38	[0.17, 0.35]	0.37***	16	[0.22, 0.52]	0.22***	23	[0.12, 0.32]	0.29***	69	[0.21, 0.36]
Analysis approach			Q_B = 1.65			Q_B = 2.52			Q_B = 0.09			Q_B = 1.42
Selected	0.50***	5	[0.26, 0.74]	0.52***	4	[0.28, 0.75]	0.30**	5	[0.07, 0.52]	0.42***	10	[0.25, 0.59]
Universal	0.33***	77	[0.27, 0.40]	0.31***	20	[0.19, 0.42]	0.26***	33	[0.16, 0.36]	0.31***	122	[0.26, 0.36]

continued on the next page...

Table 14

Methodological Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point (continued)

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Recruitment rate			Q_B = 2.80			Q_B = 2.28			Q_B = 1.26			Q_B = 0.22
Up to 50%	0.23***	25	[0.13, 0.33]	0.49***	11	[0.32, 0.66]	0.31***	15	[0.18, 0.44]	0.26***	42	[0.18, 0.33]
More than 50%	0.38***	16	[0.24, 0.51]	0.29**	8	[0.08, 0.49]	0.20**	11	[0.07, 0.34]	0.29***	29	[0.20, 0.38]
Attendance rate			Q_B = 0.07			Q_B = 1.82			Q_B = 0.01			Q_B = 1.31
Up to 50%	0.35*	4	[0.07, 0.64]	0.24*	4	[0.04, 0.43]	0.19	4	[−0.01, 0.39]	0.39***	10	[0.21, 0.56]
More than 50%	0.31***	27	[0.21, 0.42]	0.39***	13	[0.28, 0.51]	0.19***	17	[0.01, 0.28]	0.27***	50	[0.19, 0.35]
Meta-regression	0.03	31	.001	0.11	17	.012	0.04	21	.001	−0.01	60	.000
Attrition rate			Q_B = 0.31			Q_B = 0.01			Q_B = 3.36			Q_B = 1.76
Up to 20%	0.40***	64	[0.32, 0.49]	0.36***	12	[0.18, 0.53]	0.41***	18	[0.27, 0.56]	0.39***	91	[0.32, 0.46]
More than 20%	0.34***	11	[0.15, 0.53]	0.37***	11	[0.20, 0.53]	0.23**	16	[0.09, 0.37]	0.30***	30	[0.18, 0.41]
Meta-regression	−0.12	75	.02	0.15	23	.02	−0.24	34	.06	−0.13	121	.02

Notes. d = average mean difference weighted by the inverse variance, β = regression coefficient, k = number of comparisons, CI = confidence interval, Q_B = between-groups homogeneity statistics, R^2 = variance of the mean effect explained by the regression.
*$p < .05$, **$p < .01$, ***$p < .001$.

Table 15

Regression Analyses of Samples Size and an Additional Predictor on Child Outcomes

	Child training			Parent training		
Predictors	β	k	R^2	β	k	R^2
Sample size	-0.27^{**}	82	.08	-0.47^{***}	38	.22
Type of control group	-0.10			0.13		
Sample size	-0.33^{***}	82	.11	-0.40^{**}	38	.21
Study design	0.22^{*}			-0.10		
Sample size	-0.26^{**}	82	.07	-0.46^{**}	38	.20
Risk characteristics	-0.01			-0.08		
Sample size	-0.12	76	.12	-0.46^{**}	34	.22
Area of conduct	-0.26^{*}			-0.06		
Sample size	-0.24^{**}	82	.07	-0.15	38	.33
Site of study	-0.08			-0.46^{**}		
Sample size	-0.26^{**}	82	.07	-0.45^{***}	38	.23
Intensity	0.02			0.16		
Sample size	-0.26^{**}	82	.10	-0.41^{**}	38	.27
Training of the trainer	0.17			0.25		
Sample size	-0.21^{*}	82	.11	-0.69^{***}	38	.26
Implementation problems	0.22^{*}			-0.31		

Notes. β = standardised regression coefficient, k = number of comparisons, Q_{Model} = test of the model with null hypothesis that all coefficients (intercept excluded) are zero, R^2 = variance of the mean effect explained by the regression.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Sample Moderators

Among sample moderators significant and interesting results could be obtained regarding *mean age*, *proportion of dominant ethnicity*, and *proportion of low-SES* (see Table 16). More precisely, proportion of dominant ethnicity moderated the ef-

fectiveness among PTs on parental outcomes (β = 0.51, p = .028), so did mean age among CTs on child outcomes (β = -0.19, p = .026) as well as proportion of low-SES (β = 0.25, p = .003). None of the sample variables seemed to have a moderating influence among PTs on child outcomes.

Parent Trainings on Parental Outcomes (PT/PO). Regarding PT/PO, the intervention effectiveness increased with higher proportion of dominant ethnicity[18] or, in other words, with higher ethnic homogeneity (β = 0.51, p = .028). In order to understand this result, the covariance with contact between IG and CG should be considered (F[1, 25] = 6.37, p = .046). Among PTs, interventions where IG and CG had contact, the mean percentage of dominant ethnicity was higher (95.8%) than for those without contact (77.5%). More contact resulted in a descriptively higher effect size (d = 0.46, p < .001, see Table 14), compared to no contact (d = 0.26, p < .001). No significant moderation was observed regarding different dominant ethnicities ($Q_{B,all}$ = 0.35, df = 3, p = .951). Hereby, a descriptively surprisingly high effect of (d = 0.73, p < .001 could be observed for child trainings among predominantly White participants. However, this effect is based on only four comparisons and should, thus, be interpreted with caution.

Child and Parent Trainings on Child Outcomes (CT/CO and PT/CO). Among CTs the mean age of the sample had a negative influence on child outcome effect size, as the negative regression coefficient of β = -0.19 shows (p = .026). Among PTs no such prediction of child outcome effect sizes could be found, especially after removing the only comparison which conducted a PT with children older than 12 years of age, the regression coefficient strongly approximated to zero (β = -0.06, p = .675). Proportion of low-SES did significantly predict the effectiveness of CTs, as well, with higher percentage of low-SES children in the sample resulting in higher effect sizes (β = 0.25, p = .003). Among PTs no such effect was observable (β = -0.11, p = .462). Regarding covariations, both mean age and proportion of low-SES were associated with *structuredness* among CTs (mean age: F[1, 82] = 6.01, p = .016; proportion of low-SES: F[1, 82] = 10.54, p = .002). Among structured pro-

[18] A sample was considered as dominated by an ethnicity if it accounted for at least 50% of the participants.

grammes the average age was lower with M = 80.86 (SD = 41.36, k = 69) and the mean percentage of low-SES children among the sample higher with a mean of M = 91.39 (SD = 10.12, k = 69) compared to unstructured interventions (mean age: M = 113.47, SD = 56.76, k = 13; proportion of low-SES: M = 81.37, SD = 10.75, k = 13). Structured programmes, in turn, were more effective among CTs (see Table 17, Q_B = 12.62, df = 1, p < .001; d = 0.41, p < .001), compared to unstructured ones (d = 0.11, p = .143). When separately analysing the effect of age for structured versus unstructured interventions, the moderating effect of age vanished (structured: Q_B = 0.82, df = 1, p = 0.36; unstructured: Q_B = 0.90, df = 1, p = .340) and so did the predictive value of age on child outcomes when included together with structuredness into a meta-regression (β_{age} = -0.11, p = .216; $\beta_{structuredness}$ = 0.28, p = .003). The same applied to proportion of low-SES: No significant predictive influence was left among structured (p = .324) and unstructured intervention (p = .233) or when including proportion of low-SES and structuredness into one meta-regression ($\beta_{low-SES}$ = -0.15, p = .120; $\beta_{structuredness}$ = 0.25, p = .008).

The remaining variables in Table 16, that are *gender*, *dominant ethnicity*, *status of dominant ethnicity*, *additional risk factors*, and *selectivity of the sample*, did not significantly explain heterogeneity among the average effect sizes.

Two central results are subsequently summarised:

1. **PT/PO - Proportion of Dominant Ethnicity:** Higher ethnic homogeneity resulted in greater intervention effectiveness (β = 0.51). However, a higher ethnical homogeneity could be found if the intervention and control groups did have contact. That, in turn, was associated with descriptively higher effect sizes, compared to non-interacting groups. Different dominant ethnicities did not significantly moderate the mean effectiveness across both program types.

2. **CT/CO and PT/CO - Mean Age and Proportion of Low-SES:** With increasing age as well as lower proportions of low-SES children in the sample the child training effectiveness decreased (β_{age} = -0.19; $\beta_{low-SES}$ = 0.25). Hereby both moderators were associated with structuredness: Children in structured interventions were younger and more likely to be of low-SES. When controlling

for structuredness, the predictive value of mean age as well as proportion of low-SES vanished. No sample moderation could be found among parent trainings on child outcomes.

Table 16

Sample Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Mean age (in months)			Q_B = 1.64			Q_B = 1.38			Q_B = 1.02			Q_B = 0.08
Up to 24	0.45**	4	[0.11, 0.80]	0.40***	13	[0.24, 0.55]	0.27***	20	[0.14, 0.41]	0.31***	25	[0.19, 0.43]
24 to 144	0.36***	67	[0.29, 0.43]	0.33***	10	[0.17, 0.49]	0.29***	17	[0.15, 0.42]	0.33***	92	[0.27, 0.39]
Older than 144	0.25**	11	[0.08, 0.43]	0.12	1	[−0.35, 0.59]	0.00	1	[−0.54, 0.54]	0.33***	15	[0.19, 0.48]
Meta-regression	−0.19*	82	.04	−0.22	24	.05	−0.13	38	.02	−0.07	132	.01
Gender			Q_B = 1.85			Q_B = 1.24			Q_B = 1.31			Q_B = 1.50
Predominantly male	0.45	2	[−0.02, 0.919]	0.16	2	[−0.19, 0.512]	0.05	2	[−0.34, 0.434]	0.31*	5	[0.03 0.58]
Evenly distributed	0.36***	78	[0.30, 0.42]	0.37***	22	[0.26, 0.48]	0.28***	36	[0.19, 0.37]	0.33***	125	[0.28, 0.38]
Predominantly female	0.10	2	[−0.29, 0.48]	–	–	–	–	–	–	0.09	2	[−0.30, 0.47]
Meta-regression	−0.05	82	.00	0.19	24	.03	0.27	38	.07	−0.01	132	.00
Dominant ethnicity			Q_B = 5.96			Q_B = 1.36			Q_B = 1.88			Q_B = 0.35
White	0.73***	4	[0.38, 1.08]	0.23	4	[−0.07, 0.53]	0.21	6	[−0.01, 0.43]	0.35***	10	[0.14, 0.56]
African-American	0.29***	28	[0.18, 0.39]	0.42***	6	[0.22, 0.61]	0.35***	11	[0.19, 0.52]	0.31***	41	[0.22, 0.40]
Hispanic	0.33***	25	[0.22, 0.44]	0.30***	9	[0.16, 0.44]	0.26***	11	[0.11, 0.41]	0.34***	42	[0.25, 0.43]
diverse	0.37***	16	[0.24, 0.50]	0.38	1	[−0.02, 0.78]	0.15	2	[−0.14, 0.43]	0.32***	21	[0.21, 0.44]
Status of dominant ethnicity			Q_B = 0.38			Q_B = 0.58			Q_B = 0.17			Q_B = 0.07
Minority	0.28***	55	[0.22, 0.35]	0.35***	14	[0.23, 0.46]	0.27***	19	[0.15, 0.39]	0.30***	83	[0.25, 0.36]
Majority	0.35***	7	[0.15, 0.56]	0.24	5	[−0.01, 0.49]	0.32***	9	[0.13, 0.51]	0.33***	16	[0.18, 0.48]

continued on the next page...

Table 16

Sample Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point (continued)

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Proportion of dominant ethnicity			Q_B = 0.00			Q_B = 4.88*			Q_B = 0.45			Q_B = 2.01
50 to 75%	0.33***	22	[0.22, 0.43]	0.19*	6	[0.04, 0.34]	0.23**	8	[0.07, 0.40]	0.27***	33	[0.18, 0.37]
More than 75%	0.32***	34	[0.23, 0.42]	0.42***	10	[0.29, 0.56]	0.31***	17	[0.18, 0.43]	0.36***	55	[0.28, 0.45]
Meta-regression	0.08	56	.01	0.51*	16	.26	0.24	25	.06	0.24*	88	.06
Proportion of low-SES			Q_B = 8.97*			Q_B = 1.36			Q_B = 0.25			Q_B = 5.46
70 to 80%	0.23***	22	[0.11, 0.35]	0.42***	10	[0.24, 0.59]	0.28***	13	[0.14, 0.43]	0.28***	39	[0.19, 0.37]
Between 80 and 90%	0.28***	15	[0.14, 0.42]	0.20	2	[−0.15, 0.56]	0.29*	6	[0.06, 0.52]	0.25***	25	[0.13, 0.37]
More than 90%	0.45***	45	[0.36, 0.54]	0.33***	12	[0.17, 0.49]	0.24***	19	[0.12, 0.36]	0.39***	68	[0.32, 0.46]
Meta-regression	0.25**	82	.06	−0.12	24	.01	−0.11	38	.01	0.14	132	.02
Risk factors			Q_B = 0.30			Q_B = 0.05			Q_B = 0.04			Q_B = 0.45
No	0.34***	53	[0.26, 0.41]	0.34***	16	[0.20, 0.48]	0.26***	27	[0.16, 0.37]	0.31***	90	[0.25, 0.37]
Yes	0.37***	29	[0.26, 0.48]	0.37***	8	[0.18, 0.56]	0.28***	11	[0.12, 0.44]	0.35***	42	[0.26, 0.44]
Selective sample			Q_B = 0.37			Q_B = 0.20			Q_B = 1.31			Q_B = 0.32
No	0.34***	60	[0.26, 0.42]	0.35***	23	[0.23, 0.46]	0.29***	34	[0.19, 0.38]	0.32***	104	[0.26, 0.38]
Yes	0.39***	22	[0.26, 0.53]	0.49	1	[−0.13, 1.10]	0.12	4	[−0.15, 0.39]	0.36***	28	[0.24, 0.48]

Notes. d = average mean difference weighted by the inverse variance, β = standardised regression coefficient, k = number of comparisons, CI = confidence interval, Q_B = between-groups homogeneity statistics, R^2 = variance of the mean effect explained by the regression.

*$p < .05$, **$p < .01$, ***$p < .001$.

Intervention Moderators

Most of the variables tested for moderation of the average effect were characteristics of the intervention. As mentioned at the beginning of this section, a brief analysis of PT moderators on parental outcome will be presented first, followed by a more detailed investigation of possible moderators regarding the effect of CTs and PTs on child outcomes. The results are depicted in Table 17.

Parent Trainings on Parental Outcomes (PT/PO). Among PTs the effectiveness of POs was only significantly moderated by one intervention variable, that is whether implementation problems occurred or not (Q_B = 7.40, *df* = 1, *p* = .007). Contrary to the expectations, the effect size was lower if no problems occurred (d = 0.26, *p* < .001) than if problems during intervention conduct were reported (d = 0.56, *p* < .001). This result led to further analyses revealing proportion of dominant ethnicity as covarying third variable with implementation problems (*F*[1, 25] = 7.36, *p* = .012). If implementation problems occurred all samples were also ethnically more homogeneous (97.4%) compared to if no problems occurred (77.7%).

Child and Parent Trainings on Child Outcomes (CT/CO and PT/CO). Each of the following variables had a significantly moderating effect among CTs and/or PTs: *level of participation* during intervention (PT: Q_B = 4.33, *df* = 1, *p* = .038), *length of intervention* (PT: Q_B = 8.62, *df* = 2, *p* = .013), *density of the intervention* (PT: β = 0.29, *p* = .040), *application of a manual* (PT: Q_B = 14.00, *df* = 1, *p* < .001), *structuredness* (CT: Q_B = 12.62, *df* = 1, *p* < .001; PT: Q_B = 6.74, *df* = 1, *p* = .009), *implementation problems* (CT: Q_B = 6.92, *df* = 1, *p* = .009), *training of the trainer* (Q_B = 4.38, *df* = 1, *p* = .036), *supervision of the trainer* (Q_B = 5.28, *df* = 1, *p* = .022), *observation of the trainer* (Q_B = 7.29, *df* = 1, *p* = .007), *area of intervention* (CT: Q_B = 16.10, *df* = 1, *p* < .001), *social deprivation of the area* (PT: Q_B = 8.70, *df* = 1, *p* = .003), and *site of study* (PT: Q_B = 22.58, *df* = 1, *p* < .001). Only structuredness significantly moderated child outcome effectiveness among both training types.

If the participants were actively involved in the intervention (e.g., by role play) higher effect sizes could be achieved among PTs (d = 0.29, *p* < .001) as if they were only passively included (e.g., frontal knowledge transfer, d = 0.03, *p* = .818). However, the latter effect size is only based on three comparisons and will not be further investigated.

Regarding length of intervention[19], the ANOVA results revealed that short-term interventions (up to three months post intervention) were more effective (d = 0.54, p < .001) than long-term intervention (more than 12 months post, d = 0.21, p = .001) among PTs. Regarding the regression results, a significant moderation result could be found neither for CTs (β = -0.12, p = .167) nor for PTs (β = -0.28, p = .084). However, when considering the small but still significant effect sizes of interventions lasting longer than 12 months (CT: d = 0.24, p = .007; PT: d = 0.21, p = .001), it can be presumed that the regressive interdependency might not be linear but rather logarithmic. Thus, the following regressions were calculated separately for child and parent trainings:

$$Y_i = b_0 + b_1 * log(length_i) + \epsilon_i;\ i = 1, ..., k \tag{19}$$

The corresponding diagrams for CTs and PTs are displayed in Figure 6: Both models were significant with coefficients of β_1 = -0.20 (p = .027, R^2 = .04[20]) among CTs and β_1 = -0.42 (p = .007, R^2 = .18) among PTs. Furthermore, length of intervention was found to be positively correlated with attrition rate[21] (r_{CT}[75] = .30, p = .010; r_{PT}[34] = .65, p < .001). When comparing the regression coefficients of *attrition in the intervention group* versus *attrition in the control group* on child outcomes, descriptively higher values could be observed in the CG (β_{CT} = -0.15, p = .142; β_{PT} = -0.32, p = .055) than in the IG (β_{CT} = -0.09, p = .377; β_{PT} = -0.19, p = .253). The average attrition in both conditions was similar (attrition IG: M = 13.57, SD = 17.49, k = 109; attrition CG: M = 14.03, SD = 17.86, k = 108).

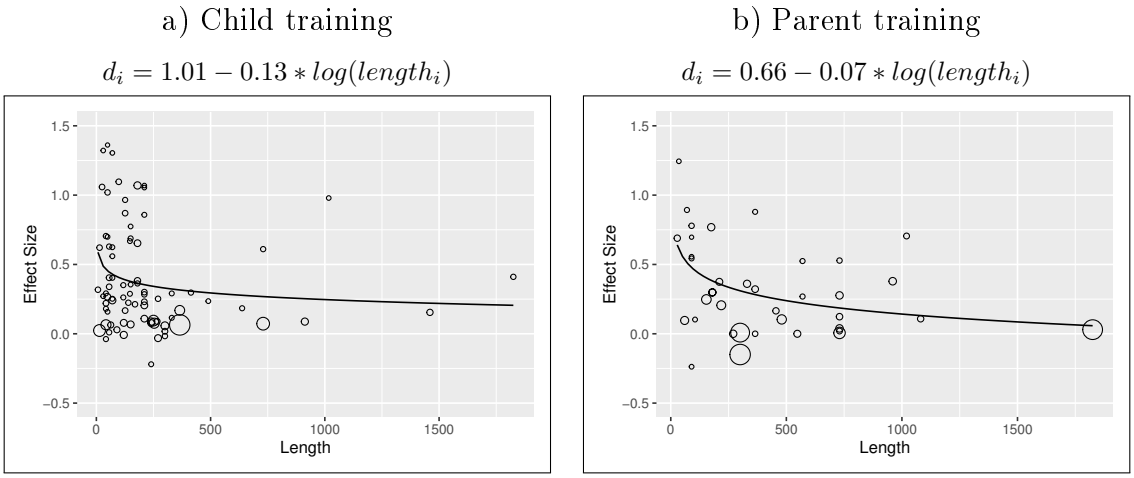
[19] Defined as time period across which the intervention has been conducted.

[20] The standardised β coefficient is reported in the text and not the unstandardised one of the regression in Figure 6 because it can be better interpreted. One outlier was excluded from the analysis.

[21] Since neither of the two variables was normally distributed a correlation according to Spearman was calculated.

Figure 6

Meta-regression of Length of Intervention on Child Outcomes at Post Time-Point



Notes. For child and parent trainings, length is being depicted on the X-axis and child outcome effect sizes at post time-point on the Y-axis, together with the logarithmic regression. Regarding the regressions, unstandardised regression coefficients are depicted.

Regarding density[22] for PTs, the effectiveness increased with higher density (β = 0.29, p = .040) but then stagnated at medium density which suggested a logarithmic interdependency rather than a linear one. Hence, a logarithmic regression of the following form was calculated:

$$Y_i = b_0 + b_1 * log(density_i) + \epsilon_i;\ i = 1, ..., k \tag{20}$$

It yielded a descriptively higher regression coefficient with β_1 = 0.33 (p = .021) and explained more variance (R^2 = .11) as can be seen in diagram b) of Figure 7. Significant mean differences were found regarding density if the area was socially deprived or not (F[1, 38] = 14.13, p < .001). Interventions realised in deprived areas had a higher mean density of 6.52 sessions per month (SD = 4.44, k = 15) as opposed to non-deprived areas with a mean of only 2.51 (SD = 2.51, k = 23). If adding social deprivation of the area to regression (20), density was not significantly different

[22]Density was calculated by dividing the overall number of sessions conducted by the length of the program in months.

from zero anymore (β = 0.16, p = .348), whereas social deprivation of the area became marginally significant in predicting the average effect size (β = 0.31, p = .061).

Among CTs, no moderating effect of density could be found (β = 0.01, p = .936). When looking at the results of the ANOVA, it is striking, though, that the mean effect sizes in the marginal groups *Low* and *High* were descriptively lower than the effect size within medium density. Hence, an inverted U-shaped curve instead of a linear regression was assumed here. Diagram a) of Figure 7 shows the density on the X-axis and child outcome effect sizes at post time-point on the Y-axis, together with the following parabolic regression[23]:

$$Y_i = b_0 + b_1 * density_i + b_2 * density_i^2 + \epsilon_i;\ i = 1, ..., k \tag{21}$$

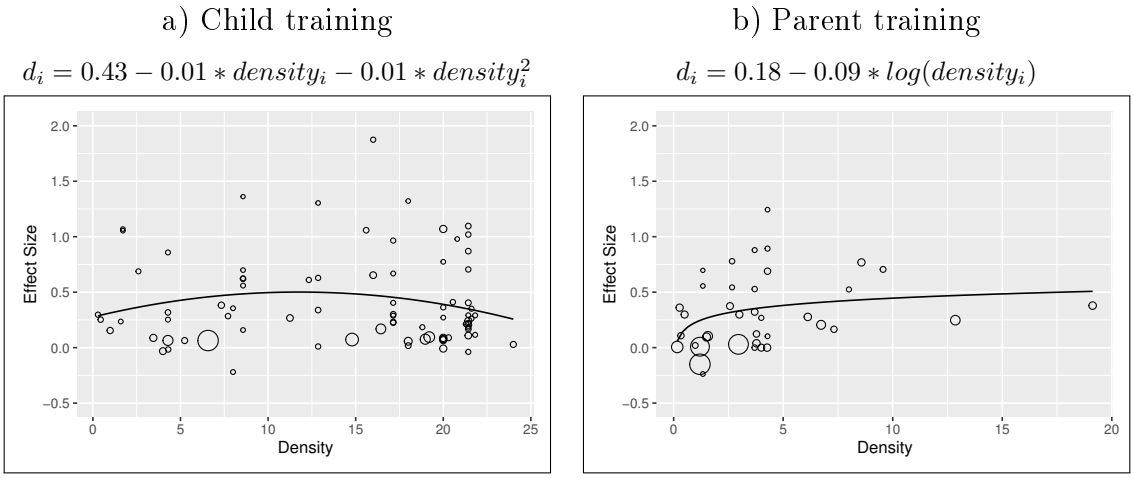
Even though the regression slope has the expected inverted U-shaped form, β_2 was not significantly different from zero (β_1 = -0.11, p = .374; β_2 = -0.16, p = .140). Further analyses showed that density was confounded with site of study (F[1, 81] = 5.71, p = .019), with multi-site studies being of higher mean density (M = 19.63, SD = 2.27, k = 9) than single-site studies (M = 13.79, SD = 7.25, k = 72). Finally, it should be mentioned that density and length of intervention were not correlated[24] (r_{CT}[82] = -.20, p = .073; r_{PT}[38] = .04, p = .827).

[23] Density was standardised in order to avoid collinearity between the predictors.

[24] A correlation according to Spearman was calculated.

Figure 7

Meta-regression of Density of the Intervention on Child Outcomes at Post Time-Point



Notes. For child and parent trainings, density is displayed on the X-axis and child outcome effect sizes at post time-point on the Y-axis, together with the parabolic (CT) and logarithmic regression (PT). Regarding the regressions, unstandardised regression coefficients are depicted.

Concerning manualisation, PT effectiveness was higher if a manual or curriculum was applied (d = 0.34, p < .001), compared to unmanualised programmes (d = 0.06, p = .361). Among CTs no such effect could be observed. Among both training types a significant association of *manualisation* and *structuredness* could be found (CT: χ^2[1, k = 82] = 6.07, p = .023; PT: χ^2[1, k = 38] = 13.46, p < .001). Hereby the ratio of an unmanualised intervention being structured was 4.62 times higher among CTs (*OR* = 0.23) than among PTs (*OR* = 0.05). Structured child and parent trainings, on the other hand, were both more effective (CT: d = 0.41, p < .001; PT: d = 0.33, p < .001) than unstructured programmes (CT: d = 0.11, p = .143; PT: d = 0.11, p = .086). In order to find out which of the two variables has a higher predictive value for child outcomes, structuredness and manualisation were both included into one meta-regression on child outcomes, separately for PTs and CTs. Among PTs, the regression coefficient of structuredness was not significantly different from zero (β = 0.10, p = .557), whereas manualisation was a substantial predictor (β = 0.40, p = .015). Among CTs, it was the other way around ($\beta_{structuredness}$ = 0.31, p < .001; β_{manual} = -0.02, p = .829).

Regarding implementation problems[25], the effectiveness was lower if problems did arise during the implementation (CT: d = 0.18; PT: d = 0.22) than if it had been conducted as planned (CT: d = 0.39; PT: d = 0.32). This moderation was not significant regarding PTs, though.

Concerning training of the trainer, if a training had been conducted with the executor the intervention was more effective among PTs (d = 0.32, p < .001) than without training (d = 0.12, p = .156). It is important to underline here that without training the effect size was not significantly different from zero (p = .156) which means that the intervention had no significant effect if the executor was untrained. In contrast to that, among CTs no significant effect of training could be observed or - quite the opposite - descriptively the effect was even reversed ($d_{untrained}$ = 0.47, p < .001 versus $d_{trained}$ = 0.34, p < .001).

It could be further observed that *supervision of the trainer* and *observation of the trainer* did significantly moderate CT effectiveness (supervision: Q_B = 5.28, *df* = 1, p = .022; observation: Q_B = 7.29, *df* = 1, p = .007). Both variables seemed to have a negative influence on child outcomes: If the trainer had been supervised (d = 0.30, p < .001) or observed (d = 0.29, p < .001) the average effect size was lower compared to no supervision (d = 0.45, p < .001) or observation (d = 0.48, p < .001). However, this effect diminished if the analysis was restricted to interventions with at least 12 months of duration (supervision: Q_B = 1.91, *df* = 1, p = .167; observation: Q_B = 1.79, *df* = 1, p = .181). Likewise, the two variables lost their moderating effect if only interventions were included which did *not* take place in a school setting (supervision: Q_B = 0.68, *df* = 1, p = .410; observation: Q_B = 1.79, *df* = 1, p = .181) or were *not* executed by educational staff (supervision: Q_B = 1.31, *df* = 1, p = .252; observation: Q_B = 2.10, *df* = 1, p = .147). In contrast to that, supervision and observation showed to have no moderating effect among PTs (supervision: d_{yes} = 0.28, p < .001 versus d_{no} = 0.26, p < .001; observation: d_{yes} = 0.33, p < .001 versus d_{no} = 0.26, p = .001).

Furthermore, implementing trainings in socially deprived areas resulted in greater

[25] An implementation problem could have been, for instance, that parts of the intervention couldn't be conducted as planned or that more than 50% of the sample dropped out before the intervention was completed.

effects (CT: d = 0.40, p < .001; PT: d = 0.42, p < .001) than in non-deprived areas (CT: d = 0.33, p < .001; PT: d = 0.17, p < .001). However, the difference did not reach significance for CT (Q_B = 1.17, df = 1, p = .280).

Effect sizes were also smaller if the intervention was conducted in rural as well as urban regions (CT: d = 0.10, p = .095; PT: d = 0.17, p = .100) and on multiple sites (CT: d = 0.23, p = .005; PT: d = 0.02, p = .787) as opposed to rural or urban areas (CT: d = 0.37, p < .001; PT: d = 0.29, p < .001) and single-site studies (CT: d = 0.38, p < .001; PT: d = 0.36, p < .001). Those two moderators were already investigated in more detail in this section in paragraph ***Methodological Moderators***.

As depicted in Table 17, the following variables did not moderate child or parent training effectiveness: *type of methods*footnoteMethods categorised as cognitive were, for instance, information transfer, psycho-education, or cognitive exercises whereas behavioural methods included role play, behavioural training, or home visitation of guided play. Accordingly, mixed methods comprised of interventions applying cognitive as well as behavioural methods., *number of methods*, *delivery format*, *individualisation*, *cultural tailoring*, *association of trainer and developer/author*, *setting of the intervention*.

In a nutshell, intervention moderator results can be summarised as follows:

1. **PT/PO - Implementation Problems:** If problems during intervention conduct occurred, higher effects could be found (d = 0.56 as if none occurred d = 0.26). If implementation problems occurred, all samples were also ethnically relatively homogeneous.

2. **CT/CO and PT/CO - Length and Density:** Regarding length of intervention, the effectiveness decreased with greater length among child trainings (β = -0.20) as well as parent trainings (β = -0.42). Length was, however, positively correlated with attrition rate. Despite a similar mean attrition in the intervention group versus control group, attrition in the control group had a descriptively stronger negative predictive effect on child outcomes than attrition in the intervention group. Regarding density among parent trainings, the effectiveness increased with higher density and stagnated at medium density at a level of d = 0.38.

When including social deprivation of the area into the regression, the regression coefficient of density was not significantly different from zero anymore. Among child trainings, a trend towards a reversely U-shaped interdependency was observed. However, there was a higher density among multi-site studies compared to single-site ones.

3. **CT/CO and PT/CO - Manulisation and Structuredness:** Manualised parent trainings were more effective (d = 0.34) than unmanualised ones (d = 0.06) which could not be observed among child trainings. In fact, unmanualised parent trainings had no significant effect on child outcomes. It was, however, much more likely to find unmanualised and structured interventions among child trainings than among parent trainings. Structured interventions, in turn, were more effective (d_{CT} = 0.41, d_{PT} = 0.33) than unstructured ones (d_{CT} = 0.11, d_{PT} = 0.11). In fact, unstructured programmes did have no significant influence on child outcomes. If including structuredness and manualisation into one meta-regression, manualisation moderated child outcomes for parent trainings, whereas structuredness among child trainings.

4. **CT/CO and PT/CO - Implementation Problems:** The effectiveness was higher among child trainings if no problems during intervention conduct occurred (d = 0.39) as opposed to with implementation problems happening (d = 0.18). No such effect could be found regarding parent trainings.

5. **CT/CO and PT/CO - Training, Supervision, and Observation:** Supervising as well as observing the trainers (d = 0.30 and 0.29, respectively) resulted in lower child training effectiveness than unsupervised and unobserved ones (d = 0.45 and 0.48, respectively). These effects diminished if only comparisons were included in the analyses which were 1.) conducted across a longer period than 12 months, 2.) not implemented in school context, or 3.) executed by non-educational staff. Training did not moderate child training effectiveness but descriptively higher effects were observed for untrained compared to trained executors. Among parent trainings, interventions with trained executors were of small effect size (d = 0.32) whereas untrained ones had no statistically significant effect (d = 0.12). No moderating effect could be observed concerning

supervision and observation among parent trainings.

6. **Common CT/CO and PT/CO - Social Deprivation of the Area:** Parent training effectiveness was higher if the programmes were conducted in socially deprived areas (d = 0.42) as opposed to in non-deprived areas (d = 0.17). No such moderation was found among child trainings.

Table 17

Intervention Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Type of methods			Q_B = 2.14			Q_B = 0.04			Q_B = 0.60			Q_B = 1.36
Cognitive	0.42***	31	[0.31, 0.529]	0.39	2	[−0.02, 0.80]	0.18	4	[−0.05, 0.42]	0.37***	37	[0.28, 0.47]
Behavioural	0.28**	9	[0.07, 0.49]	0.34*	3	[0.03, 0.65]	0.26**	9	[0.08, 0.44]	0.28***	19	[0.14, 0.42]
Mixed	0.32***	42	[0.23, 0.42]	0.35***	19	[0.22, 0.48]	0.28***	25	[0.17, 0.39]	0.32***	75	[0.25, 0.39]
Number of methods			Q_B = 3.17			Q_B = 1.28			Q_B = 1.79			Q_B = 0.95
One to two	0.41***	45	[0.32, 0.50]	0.50*	2	[−0.12, 0.88]	0.14	6	[−0.05, 0.34]	0.35***	56	[0.27, 0.43]
Three to four	0.28***	29	[0.17, 0.39]	0.37***	15	[0.23, 0.52]	0.28***	24	[0.17, 0.39]	0.30***	56	[0.22, 0.37]
Five and more	0.34**	8	[0.11, 0.56]	0.27**	7	[0.08, 0.47]	0.31***	8	[0.13, 0.49]	0.34***	19	[0.20, 0.47]
Meta-regression	−0.15	82	.02	−0.24	24	.06	0.23	38	.05	−0.04	131	.00
Delivery format			Q_B = 1.77			Q_B = 2.89			Q_B = 3.63			Q_B = 1.39
Individual	0.31***	15	[0.15, 0.46]	0.29***	13	[0.14, 0.45]	0.26***	23	[0.15, 0.37]	0.28***	39	[0.19, 0.38]
Group	0.38***	59	[0.30, 0.46]	0.20	2	[−0.13, 0.53]	0.13	6	[−0.06, 0.31]	0.34***	68	[0.26, 0.41]
Both	0.25*	8	[0.06, 0.46]	0.46***	9	[0.29, 0.64]	0.38***	9	[0.20, 0.55]	0.37***	24	[0.26, 0.49]
Individualisation			Q_B = 0.22			Q_B = 2.32			Q_B = 0.04			Q_B = 0.00
Low	0.33***	27	[0.21, 0.44]	0.44***	11	[0.28, 0.59]	0.28***	15	[0.14, 0.43]	0.33***	48	[0.25, 0.42]
High	0.36***	55	[0.28, 0.44]	0.27***	13	[0.13, 0.42]	0.26***	23	[0.14, 0.38]	0.33***	82	[0.26, 0.39]
Level of participation			Q_B = 0.07			Q_B = 0.01			Q_B = 4.33*			Q_B = 1.18
Low	0.33**	9	[0.13, 0.52]	0.38	1	[−0.11, 0.87]	0.03	3	[−0.21, 0.26]	0.25***	15	[0.11, 0.40]
High	0.35***	73	[0.28, 0.42]	0.35***	23	[0.24, 0.46]	0.29***	35	[0.20, 0.38]	0.34***	116	[0.28, 0.39]

continued on the next page...

Table 17

Intervention Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point (continued)

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Length of intervention (in months)			Q_B = 2.55			Q_B = 2.90			Q_B = 8.62*			Q_B = 6.56*
Up to three	0.41***	29	[0.29, 0.53]	0.41***	7	[0.18, 0.64]	0.54***	9	[0.34, 0.75]	0.43***	41	[0.33, 0.53]
Three to 12	0.35***	42	[0.26, 0.44]	0.24**	8	[0.09, 0.40]	0.20**	11	[0.06, 0.35]	0.30***	59	[0.23, 0.38]
More than 12	0.24**	11	[0.07, 0.41]	0.43***	9	[0.26, 0.61]	0.21**	18	[0.08, 0.33]	0.26***	32	[0.16, 0.36]
Meta-regression	−0.12	82	.01	0.18	24	.03	−0.28	38	.08	−0.16*	131	.02
Density of intervention[a]			Q_B = 1.29			Q_B = 1.71			Q_B = 2.49			Q_B = 3.77
Low (up to 6)	0.29***	15	[0.12, 0.46]	0.33***	17	[0.20, 0.47]	0.22***	30	[0.13, 0.32]	0.27***	52	[0.19, 0.36]
Medium (6 to 13)	0.43***	17	[0.26, 0.60]	0.34***	6	[0.14, 0.53]	0.38***	7	[0.20, 0.57]	0.39***	28	[0.27, 0.51]
High (more than 13)	0.37***	49	[0.28, 0.46]	0.67**	1	[0.18, 1.15]	0.38	1	[−0.10, 0.86]	0.37***	50	[0.29, 0.46]
Meta-regression	0.01	81	.00	0.21	24	.04	0.29*	38	.08	−0.08	131	.01
Cultural tailoring			Q_B = 0.47			Q_B = 0.39			Q_B = 1.74			Q_B = 0.58
No	0.36***	64	[0.29, 0.44]	0.31***	11	[0.15, 0.48]	0.21***	22	[0.11, 0.32]	0.31***	92	[0.25, 0.37]
Yes	0.31***	18	[0.18, 0.44]	0.38***	13	[0.24, 0.53]	0.33***	16	[0.20, 0.46]	0.36***	38	[0.27, 0.45]
Manualisation			Q_B = 0.09			Q_B = 0.01			Q_B = 14.00***			Q_B = 3.81
No	0.33***	17	[0.19, 0.48]	0.36**	6	[0.12, 0.61]	0.06	11	[−0.06, 0.18]	0.24***	30	[0.13, 0.34]
Yes	0.36***	65	[0.28, 0.43]	0.35***	18	[0.22, 0.47]	0.34***	27	[0.25, 0.43]	0.35***	100	[0.30, 0.41]
Structuredness			Q_B = 12.62***			Q_B = 0.01			Q_B = 6.74**			Q_B = 16.78***
Low	0.11	13	[−0.04, 0.26]	0.36**	8	[0.14, 0.58]	0.11	14	[−0.02, 0.24]	0.14**	29	[0.04, 0.24]
High	0.41***	69	[0.34, 0.48]	0.35***	16	[0.22, 0.48]	0.33***	24	[0.23, 0.42]	0.38***	101	[0.33, 0.44]

continued on the next page...

Table 17

Intervention Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point (continued)

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Training of the trainer			Q_B = 2.24			Q_B = 0.82			Q_B = 4.38*			Q_B = 0.09
No	0.47***	20	[0.33, 0.62]	0.27*	6	[0.05, 0.48]	0.12	10	[−0.05, 0.28]	0.33***	34	[0.22, 0.43]
Yes	0.34***	57	[0.26, 0.42]	0.38***	18	[0.25, 0.51]	0.32***	28	[0.22, 0.42]	0.35***	91	[0.28, 0.41]
Supervision of the trainer			Q_B = 5.28*			Q_B = 1.67			Q_B = 0.07			Q_B = 0.27
No	0.45***	40	[0.35, 0.55]	0.29***	13	[0.14, 0.43]	0.26***	17	[0.12, 0.39]	0.36***	62	[0.28, 0.44]
Yes	0.30***	38	[0.20, 0.39]	0.43***	11	[0.27, 0.58]	0.28***	21	[0.16, 0.41]	0.33***	64	[0.25, 0.40]
Observation of the trainer			Q_B = 7.29**			Q_B = 3.07			Q_B =0.47			Q_B = 2.35
No	0.48***	37	[0.38, 0.59]	0.43***	14	[0.29, 0.57]	0.26***	21	[0.13, 0.39]	0.39***	62	[0.31, 0.47]
Yes	0.29***	41	[0.19, 0.38]	0.23*	9	[0.04, 0.42]	0.33***	15	[0.18, 0.49]	0.31***	61	[0.23, 0.38]
Implementation problems			Q_B = 6.92**			Q_B = 7.40**			Q_B = 0.78			Q_B = 6.45*
No	0.39***	68	[0.32, 0.47]	0.26***	17	[0.16, 0.37]	0.32***	26	[0.20, 0.43]	0.37***	102	[0.31, 0.43]
Yes	0.18*	14	[0.03, 0.32]	0.56***	7	[0.38, 0.74]	0.22*	11	[0.04, 0.40]	0.21***	28	[0.10, 0.32]
Association of trainer and..			Q_B = 2.99			Q_B = 0.87			Q_B = 2.99			Q_B = 3.65
developer	0.27**	16	[0.08, 0.47]	0.18	2	[−0.22, 0.58]	0.04	3	[−0.34, 0.41]	0.24**	21	[0.09, 0.39]
author	0.51***	16	[0.32, 0.70]	0.39***	11	[0.20, 0.58]	0.40***	15	[0.23, 0.58]	0.42***	34	[0.31, 0.54]
Setting			Q_B = 2.22			Q_B = 1.16			Q_B = 4.88			Q_B = 2.70
Educational facility	0.36***	71	[0.29, 0.42]	0.20	2	[−0.15, 0.55]	0.12	4	[−0.09, 0.32]	0.33***	82	[0.27, 0.39]
Home	0.12	2	[−0.28, 0.51]	0.33***	10	[0.14, 0.52]	0.24***	20	[0.13, 0.36]	0.25***	22	[0.12, 0.37]
Several	0.51**	3	[0.14, 0.90]	0.400***	7	[0.18, 0.61]	0.31**	7	[0.13, 0.51]	0.400***	14	[0.24, 0.55]
Other	0.30*	6	[0.06, 0.55]	0.41***	5	[0.17, 0.65]	0.44***	6	[0.23, 0.66]	0.38***	12	[0.21, 0.55]

continued on the next page...

Table 17

Intervention Moderator Analyses (ANOVAs and Meta-Regressions) by Type of Intervention and Outcome Categories at Post Time-Point (continued)

	Child training			Parent training								
	Child outcomes			Parental outcomes			Child outcomes			Overall outcomes		
Moderator	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2	d/β	k	95% CI/R^2
Area of intervention			Q_B = 16.10***			Q_B = 0.52			Q_B = 1.10			Q_B = 10.86**
Rural or urban	0.37***	65	[0.31, 0.44]	0.38***	15	[0.24, 0.52]	0.29***	26	[0.18, 0.40]	0.35***	100	[0.29, 0.41]
Rural and urban	0.10	11	[-0.02, 0.22]	0.28*	6	[0.04, 0.52]	0.17	8	[-0.03, 0.37]	0.15**	22	[0.05, 0.26]
Social deprivation of the area			Q_B = 1.17			Q_B = 2.97			Q_B = 8.70**			Q_B = 7.32**
No	0.33***	57	[0.25, 0.40]	0.27***	14	[0.13, 0.41]	0.17***	23	[0.07, 0.27]	0.28***	90	[0.22, 0.34]
Yes	0.40***	25	[0.29, 0.51]	0.46***	10	[0.30, 0.62]	0.42***	15	[0.29, 0.56]	0.43***	42	[0.34, 0.54
Site of study			Q_B = 2.59			Q_B = 1.03			Q_B = 22.58***			Q_B = 19.67***
Single-site	0.38***	72	[0.30, 0.45]	0.37***	22	[0.25, 0.49]	0.36***	30	[0.27, 0.44]	0.37***	105	[0.32, 0.42]
Multi-site	0.23**	10	[0.07, 0.39]	0.18	2	[−0.17, 0.53]	0.02	8	[−0.100, 0.13]	0.13**	25	[0.05, 0.22]

Notes. d = average mean difference weighted by the inverse variance, β = standardised regression coefficient, k = number of comparisons, CI = confidence interval, Q_B = between-groups homogeneity statistics, R^2 = variance of the mean effect explained by the regression.

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. ^{a}number of sessions per month. Among CT and overall trainings one outlier has been excluded lying about nine standard deviations above the mean.

3.2.6 Differentiated Outcome-Analyses

As described in section 2.1.4, a finer effectiveness distinction regarding child outcomes was intended to be conducted by analysing programme effectiveness on basic development (BD) and school development (SD). Hereby, it will be differentiated between child and parent training effectiveness. It is important to mention at this point that the composition regarding descriptive characteristics of the sample and intervention were different if all child outcomes were analysed as opposed to if only BD or SD were considered. Regarding child and parent trainings, the mean age of the samples when analysing BD outcomes was much younger (4 years and 10 months as well as 1 year and 6 months for CTs and PTs, respectively) as opposed to when focussing on SD (7 years as well as 3 years and 6 month, respectively). Furthermore, among CTs the method *free or guided play* was 10.4% more likely to be applied if BD versus SD was analysed. Regarding the method *performances* it was 15.5%.

Short-term Effectiveness. The average CT effect regarding BD was small with $d = 0.42$ ($SE = 0.02$) but significantly different from zero ($p < .001$). It was descriptively higher than the average CT/CO effect ($d = 0.35$, $p < .001$). There was substantial true variance among the effect sizes ($I^2 = 85.5\%$, $p < .001$). Regarding SD, a descriptively lower average effect than CT/CO could be observed. It was of small size with $d = 0.32$ ($SE = 0.03$) and statistically significant ($p < .001$). Heterogeneity analysis revealed substantial true variance regarding the effect sizes ($I^2 = 82.6\%$, $p < .001$).

The mean PT effect on BD was small and significant ($d = 0.32$, $SE = 0.06$, $p < .001$) and the effect sizes heterogeneous ($I^2 = 69.2\%$, $p < .001$). The effect was descriptively higher compared to PT/CO effectiveness ($d = 0.27$, $p < .001$). The average SD effectiveness was descriptively lower, but significant ($d = 0.20$, $SE = 0.05$, $p < .001$). Again, significant heterogeneity among the effect sizes was present ($I^2 = 80.4\%$, $p < .001$). Long-term effectiveness analyses were conducted, but the case number was too low to obtain reliable results. Thus, the findings were deliberately not displayed.

Since each of the average effect sizes did contain true variance, selected moderator analyses were conducted. The variables were chosen based on their relevancy which was determined by their predictive value on outcome category level and their reliability which was evaluated based on a minimal number of five comparisons in each cell. Since the moderator analyses were conducted separately for CTs and PTs, differ-

ent moderators were omitted for each of the two intervention types. Regarding CT analyses, for instance, manualisation was excluded and structuredness included, as regression analyses on child outcomes suggested (see section 3.2.5, paragraph ***Intervention Moderators***). Following this reasoning, the selection was reversed for PTs. Furthermore, it was considered sufficient investigating the influence of methodological moderators on outcome category level. The detailed moderating results can be obtained from Table 18 regarding CTs and from Table 19 concerning PTs.

Moderator Analyses for Child Trainings. Structuredness significantly moderated school development effectiveness (Q_B = 13.01, *df* = 1, $p < .001$), yielding a significant effect size for structured programmes (d = 0.38, $p < .001$) and no substantial effect when interventions were not structured (d = 0.09, $p = .222$). Regarding basic development, no moderation could be found (Q_B = 0.20, *df* = 1, $p = .656$). However, only four comparisons did conduct an unstructured intervention. Supervising and observing the trainer resulted in lower school development effect sizes (supervision: d = 0.29, $p < .001$; observation: d = 0.28, $p < .001$) than unsupervised (d = 0.41, $p < .001$) and unobserved programmes (d = 0.44, $p < .001$). The moderation was substantial for observation (Q_B = 5.23, *df* = 1, $p = .022$), but only marginally reached significance regarding supervision (Q_B = 2.76, *df* = 1, $p = .097$). No moderating influence could be found regarding BD outcomes, neither for supervision (Q_B = 0.22, *df* = 1, $p = .639$), nor for observation (Q_B = 0.52, *df* = 1, $p = .472$). Regarding implementation problems, lower average effect sizes could be revealed for BD and SD, if problems during programme conduct happened (BD: d = 0.18, $p = .344$; SD: d = 0.18, $p = .016$), than if none occurred (BD: d = 0.46, $p < .001$; SD: d = 0.36, $p < .001$). The difference was substantial among SD (Q_B = 5.33, *df* = 1, $p = .021$), but not among BD (Q_B = 2.02, *df* = 1, $p = .155$). It can be expected, though, that a higher number of interventions reporting implementation problems would yield a significant moderation.

Moderator Analyses for Parent Trainings. The moderating effect of manualisation was confirmed also regarding BD (Q_B = 17.15, *df* = 1, $p = .008$) and SD (Q_B = 14.76, *df* = 1, $p < .001$). If a manual was provided, the average effect size was small and significant from zero (BD: d = 0.39, $p < .001$; SD: d = 0.34, $p < .001$), whereas if no manual was used there was no significant average effect neither on BD (d = 0.11, $p = .236$), nor on SD (d = 0.00, $p = .966$). A similar picture emerged regarding train-

ing of the trainer: Small, significant average effect sizes if a training was conducted (BD: d = 0.35, p < .001; SD: d = 0.31, p < .001) versus no significant results if it wasn't (BD: d = 0.22, p = .066; SD: d = 0.06, p = .531). However, training only reached a significant moderating influence for SD (Q_B = 4.50, df = 1, p = .034), but not for BD (Q_B = 0.83, df = 1, p = .362). Social deprivation of the area was found to be a moderator regarding BD (Q_B = 17.57, df = 1, p < .001), but not SD (Q_B = 2.03, df = 1, p = .154), in the direction that if interventions were conducted in deprived areas higher effect sizes could be observed (d_{BD} = 0.55, p < .001) than if the areas were not deprived (d_{BD} = 0.16, p = .004).

The following statement sum up the results regarding the differentiated outcome construct analyses.

1. **Descriptive Characteristics:** The samples regarding both training types were about 2 years younger when basic development was investigated, compared to school development. Furthermore, free or guided play and performance as methods were more often applied among basic development, compared to school development among child trainings.

2. **Average Effect:** All average short-term effects were of small size (ranging between d = 0.23 and 0.42), but significantly different from zero. All contained true variance among the effect sizes. Child trainings were descriptively more effective regarding basic and school development than parent trainings.

3. **Child Training Moderators:** None of the analysed variables moderated basic development effectiveness. Only regarding implementation problems, a trend could be found in the direction that if problems occurred, the average effect size was lower (d = 0.18) as if none were reported (d = 0.46). Regarding school development, structured, unsupervised, unobserved, and well-implemented interventions yielded higher effect sizes (ranging between d = 0.36 and 0.44) than unstructured, supervised, observed, and poorly implemented ones (ranging from d = 0.09 to 0.29). Concerning supervision, though, this difference needs to be regarded as a trend.

4. **Parent Training Moderators:** If no manual was provided, parent trainings had no significant effect on basic or school development, whereas manual application led to small and significant average effects (d_{BD} = 0.39; d_{SD} = 0.34). Training showed to be significantly influencing school development since higher intervention effectiveness emerged for trained than untrained interventions (d = 0.31 and 0.06, respectively). That was not the case regarding basic development. Moreover, interventions implemented in socially deprived areas resulted in a higher average basic development effect (d = 0.55) compared to if conducted in non-deprived areas (d = 0.39). No such influence could be found for school development regarding the social deprivation of the area.

Table 18

Differentiated Child Outcome Construct Analyses Regarding Selected Intervention Moderators for Child Trainings at Post Time-Point

	Basic development			School development		
Moderator	d	k	95% CI	d	k	95% CI
Structuredness			$Q_B = 0.20$			$Q_B = 13.01^{***}$
Low	0.35	4	[−0.02, 0.71]	0.09	13	[−0.05, 0.23]
High	0.43***	29	[0.29, 0.57]	0.38***	64	[0.31, 0.46]
Supervision of the trainer			$Q_B = 0.22$			$Q_B = 2.76$
No	0.48***	15	[0.27, 0.69]	0.41***	37	[0.30, 0.51]
Yes	0.41***	17	[0.23, 0.60]	0.29***	36	[0.20, 0.38]
Observation of the trainer			$Q_B = 0.52$			$Q_B = 5.23^{*}$
No	0.50***	16	[0.30, 0.70]	0.44***	33	[0.33, 0.55]
Yes	0.39***	16	[0.20, 0.59]	0.28***	40	[0.18, 0.37]
Implementation problems			$Q_B = 2.02$			$Q_B = 5.33^{*}$
No	0.46***	29	[0.32, 0.60]	0.36***	63	[0.29, 0.44]
Yes	0.18	4	[−0.19 0.54]	0.18*	14	[0.03, 0.32]

Notes. d = average mean difference weighted by the inverse variance, k = number of comparisons, CI = confidence interval, Q_B = between-groups homogeneity statistics.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 19

Differentiated Child Outcome Construct Analyses Regarding Selected Intervention Moderators for Parent Trainings at Post Time-Point

	Basic development			School development		
Moderator	d	k	95% CI	d	k	95% CI/
Manualisation			$Q_B = 7.15^{**}$			$Q_B = 14.76^{***}$
No	0.11	8	[−0.07, 0.28]	0.00	6	[−0.13, 0.13]
Yes	0.39***	20	[0.28, 0.51]	0.34***	18	[0.23, 0.45]
Training of the trainer			$Q_B = 0.83$			$Q_B = 4.50^{*}$
No	0.22	7	[−0.02, 0.46]	0.06	6	[−0.13, 0.26]
Yes	0.35***	21	[0.22, 0.48]	0.31***	18	[0.18, 0.45]
Social deprivation of the area			$Q_B = 17.57^{***}$			$Q_B = 2.03$
No	0.16**	16	[0.05, 0.26]	0.18**	14	[0.05, 0.30]
Yes	0.55***	12	[0.40, 0.70]	0.33***	10	[0.16, 0.51]

Notes. d = average mean difference weighted by the inverse variance, k = number of comparisons, CI = confidence interval, Q_B = between-groups homogeneity statistics.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

4 Discussion

The goal of this meta-analysis was to summarise the high quality research evidence regarding psychosocial and educational interventions which aim at preventing educational consequences of poverty for children and adolescents worldwide. The retrieved studies go back to 1967. The most recent one included was published in 2013. Based on defined eligibility criteria, 109 publications, yielding 132 comparisons between an intervention and an untreated control group, could here be investigated. Most of the studies were implemented in the United States and published in peer-reviewed journals. The findings are based on almost 1 300 effect measures and more than 80 000 children and adolescents. The children are predominantly between 2 and 12 years of age and mainly have an African-American or Hispanic ethnic background. The trainings typically targeted cognitive development, language skills, or learning skills, working either directly with the child, or indirectly by approaching its parents. If children were the intervention subject, the most common mode of intervention were exercises. If parents were addressed, home visits and/or behavioural trainings were typically implemented. Most programmes were realised in urban school settings.

Overall, the studies included in this meta-analysis can be regarded as solid evidence and representative, providing data that allows to answer the research questions of the dissertation. Furthermore, the findings contribute to clarify the precipitating mechanisms regarding intervention effectiveness.

On the following pages the findings will be reflected, beginning with the overall effectiveness and its stability (section 4.1). Next, the moderation of the effectiveness by several relevant variables will be discussed (section 4.2), briefly for all parental outcomes (section 4.2.1) and, in more details, regarding child development outcomes (sections 4.2.2 and 4.2.3). Then, recommendations for research and practise will be given (sections 4.3 and 4.4), followed by a discussion regarding the limitations of the meta-analysis (section 4.5). Last but not least, a final conclusion will sum up this dissertation's findings (section 4.6).

4.1 Effectiveness of Preventive Interventions

As mentioned in section 4.1.1, meta-analyses are at risk of finding distorted results because of selective study publication (Pigott, 2009; Rosenthal, 1979; Sterling, 1959). Hence, before evaluating the overall and long-term findings of preventive interventions, their distortion by publication bias needs to be addressed. The typical interpretation of sample size and effect size association (i.e., selective publication processes), will be questioned in this context and other factors will be introduced as possible explanations. Subsequently, the observed overall effects, as well as the short-term effects by training type, and the long-term effects will be related to similar meta-analyses in this field (see sections 4.1.2, 4.1.3, and 4.1.4). Furthermore, the observed overall effects will be discussed in relation to the national norm (see section 4.1.2, paragraph ***Comparison to the National Norm***).

4.1.1 Publication Bias

Even though the probability for a publication bias has been minimised by including only studies with a minimal sample size of 50, all tests for publication bias were positive, even if only very large studies were analysed. Since large publications should not be biased by selective publication processes (see section 2.6), those findings led to the conclusion that publication bias was not the cause for the distortion of the results but rather other variables that covaried with sample size (Beelmann, 2006; Beelmann, Pfost, & Schmitt, 2014; Egger et al., 1997).

As the findings show sample size ceased to be a predictor for the average effect size if area of conduct among child trainings and site of study among parent trainings were included as additional variables into the meta-regressions. In order to understand the mechanisms behind those variables, covarying factors concerning these two implementation characteristics of the intervention were analysed. For child trainings, that led to the identification of supervision as a covarying factor. Hereby, intervention implemented in mixed areas were also more likely to be supervised than programmes conducted in rural or urban areas only. Supervision, on the other hand, had a negative influence on child development. This finding will be discussed in detail further below (see section 4.2.2, paragraph ***Training, Supervision, and Obser-***

vation). Among parent trainings, more manualised programmes and interventions conducted in socially deprived areas were present among single-site studies, compared to multi-site ones. Manualised interventions, in turn, were more effective than unmanualised ones and so were interventions implemented in deprived as opposed to non-deprived areas. Those two moderators will be discussed further in section 4.2.2, paragraph ***Structuredness and Manualisation*** and ***Social Deprivation of the Area***. It is possible that other factors were involved which were not assessed here. Studies conducted across urban and rural areas as well as studies implemented in multiple sites have in common that usually many centres are involved. That could be associated with a number of different problems. It can be assumed, for instance, that the geographically wide spread centres usually operate independently of each other. Furthermore, for a larger-scale study or a study which is conducted in mixed areas, a high number of centres is necessary for obtaining a representative sample. Thus, it is probable that centres had to be included which didn't fit the scientific standard or are inexperienced in conducting an externally introduced intervention. There might have been, for instance, flaws in the randomisation process which have not been mentioned to the study authors or problems in implementing the intervention as planned. Consequently, a higher variation in the process of intervention conduct, participant retention, or outcome assessment can be expected. This, in turn, might have resulted in larger measurement errors which decreased the observed effect size and by that the overall effect size, since higher weights were assigned to large studies when calculating the overall effect size (Rothstein, Sutton, & Borenstein, 2005). Independently of the experience centres have in conducting intervention studies, there are further aspects which might influence the effectiveness of large trials. They "...require extensive coordination, funding, and infrastructure..." (Rothstein et al., 2005, p. 294), all factors which are "... hard to control and virtually inevitable, especially in field experiments" (Beelmann & Lipsey, in press, p. 2). Furthermore, those variables are difficult to measure and are hardly ever reported in research articles. Another possible influential variable could have been the relationship between the trainer and the trainees - a variable extensive to assess and easily affected by social desirability, especially in school contexts. Not a single study included in this meta-analysis did evaluate the trainer-trainee-relationship.

In a nutshell, a distortion of the effects by selective publication processes can be ruled out because other factors that covary with sample size mediated the prediction of sample size on effect size. For child trainings the interdependency of sample size and effect size can be explained by the association of sample size with area of conduct which in turn was associated with supervision: Higher samples sizes are more likely to be found in interventions conducted in mixed areas and are also more likely to be supervised. For parent trainings the explanatory variable turned out to be site of study which was confounded with manualisation and social deprivation of the area. It can be concluded that future research should collect data on variables which could be associated with sample size and effect size. In doing so, the phenomenon could be analysed more differentiated and by that, the relative influence of publication bias, implementation variables, and/or trainer-trainee-relationship on the average effect size can be unravelled.

4.1.2 Overall Effectiveness of Preventive Interventions

Across all comparisons and time-points a small but significant average intervention effect of textitd = 0.31 was observed. This equals a success rate of 15% in the intervention compared to the control group. This result is consistent with findings from comparable research regarding low-SES children and adolescents. Kim and Quinn (2013) analysed the effectiveness of summer reading programmes on reading achievement and found in their subgroup analysis of low-income samples an average effect size of $d = 0.33$. Slightly lower but also significantly different from zero was the effectiveness of centre-based early education interventions on cognitive and school outcomes synthesised by Camilli et al. (2010) with $d = 0.19$. However, not all samples in this meta-analysis had a low socio-economic background which might explain a lower effect size. Furthermore, Burger (2010) reviewed early childhood care and education interventions and judged the findings on cognitive achievement as beneficial. Manning et al. (2010) found a slightly higher effect of $d = 0.43$[26]. It should be considered, though, that they strongly selected a certain kind of studies by only including those which

[26] Retrieved by calculating the mean of the outcome domains *educational success* and *cognitive development*.

measured one follow-up during adolescence in addition to a post assessment. It can be assumed that this inclusion criterion led to the selection of particularly effective interventions because studies which did not find short-term effects, might have not even evaluated the development throughout adolescence.

The sensitivity analysis showed that the findings of this dissertation were robust to alternative decisions concerning the publication format, the coder of the reports, the age of the sample, and the type of low-SES. Thus, the findings can be regarded with a higher degree of certainty.

Comparison to the National Norm. By comparing the educational development of the intervention group to the national representative norms it would possible to estimate the extent of compensation by psychosocial and educational interventions regarding the educational developmental delay of children and adolescents of low-SES. Due to the lack of reports on age-standardised, norm-referred outcome measures, though, no such analysis could be conducted in this meta-analysis. However, comprehensive and representative surveys which estimated this compensation are the *Head Start Family* and *Child Experiences Survey FACES* (Zill et al., 2006), the *Head Start Impact Study* (Puma, Bell, Cook, Heid, & Lopez, 2005), and the *Miami School Readiness Project* (Winsler et al., 2008). In 1997 FACES started to investigate the cognitive and social-emotional development of children by drawing nationally representative cohorts of Head Start programmes and compared the observed effects with norming samples. The age of the children at programme entry was 3-4 years and post assessments were conducted after 1-2 years of intervention conduct. The Head Start Impact Study was included as primary study into this meta-analysis and is conceptually equivalent to FACES, thus, the results are comparable. Overall, gains could be observed regarding vocabulary, early math, or early writing outcomes. Yet, the children still have a long way to go in order to meet the average performance level of U.S. children. Head Start graduates (defined as those who graduated after two years of Head Start) did gain most by between one-quarter to three-quarters of a standard deviation which corresponds to about 12 to 35%, depending on the outcome measured (vocabulary, early math, or early writing). The greatest convergence to the national norm could be observed for pre-reading skills. Precisely, Head Start graduates were only about one third of a standard deviation below the national norm when they

entered first grade which is equivalent to a difference of only 16%. However, Head Start graduates considerably lacked behind the norm in vocabulary and early math. A much more promising approach seem to be pre-kindergarten programmes (i.e., for low-SES children, those are the groups *subsidised centre-based childcare* and *free Title I public school pre-k programmes*[27]), as Winsler et al. (2008) concluded after analysing data from the research and evaluation project Miami School Readiness. The children were assessed at the beginning and at the end of pre-kindergarten and, again, considerable gains could be achieved: In all three measured domains (i.e., cognitive, language, and fine motor development) they passed the national average mark by the end of the school year (57th, 52nd, and 58th percentile, respectively). Unfortunately, those results were not contrasted to any kind of untreated group. It would have further been interesting to analyse the stability of those effects by a later follow-up assessment.

4.1.3 Short-Term Effectiveness by Intervention Type

Child and parent trainings (the latter commonly conducted with the mother) showed similar short-term effects on a small level across all outcomes as well as regarding child outcomes ($d = 0.35$ and 0.27, respectively). The findings regarding child trainings are in accordance with those of other comparable and recent meta-analyses (Burger, 2010; Camilli et al., 2010; Darrow, 2009; Kim & Quinn, 2013) and so is the overall short-term influence of parent trainings (Astuto & Allen, 2009; Brooks-Gunn & Markman, 2005). Surprisingly high, though, was the parent training effect on child outcomes observed in this dissertation since other research found little to no evidence of a parent to child transmission (Astuto & Allen, 2009; Brooks-Gunn & Markman, 2005). Maybe, this is due to the selective and relatively homogeneous low-SES sample in this meta-analysis (see section 3.1.2, paragraph ***Sample Characteristics***). Support comes from Sweet and Appelbaum (2004), who observed that home visitation was more effective among low-income groups as opposed to such studies which didn't target low-income samples. If considering the moderate correlation of

[27] Title I is a programme which financially supports schools with high percentages of low-incomes children.

income and education (Bollen, Glanville, & Stecklov, 2001; Hauser & Huang, 1997), it is reasonable to assume that the effect among low-SES samples might be higher because children from low-income backgrounds are also more likely to be educationally disadvantaged. Hence, the potential for improvement is higher, both for children and for parents.

Overall teacher training effectiveness was lower but still substantial ($d = 0.18$). It needs to be considered, though, that six of the seven comparisons were cluster-randomised with a mean sample size of 1 992. As already discussed in section 4.1.1, larger-scale studies are more difficult to implement due to higher complexity regarding, for instance, programme distribution to geographically wide-spread study sites or because of more extensive participant retention. What surprises, though, is that the indirect effect of teacher trainings on child outcomes is hardly noticeable. However, it still equals to a 3% average improvement rate compared to the untreated groups. Considering the high sphere of influence by each teacher due to an average primary school class size of 21 in the OECD (OECD, 2015), teacher trainings change little, but little in many recipients which might be an option for an efficient approach. Unfortunately, the number of studies conducted in this area are still too few to further investigate influential variables.

The effectiveness of a combination of child and parent trainings was highest for all training types regarding overall as well as child outcomes ($d = 0.45$ and $d = 0.44$, respectively), but didn't reach significance. That is due to the low number of five comparisons which actually implemented combined trainings, but also to the strong variability among the effect sizes, ranging from 0.13 to 1.24. Manning et al. (2010) investigated the moderating effect of, as they characterised it, *number of components*[28]. Their formed category *one or two components* is comparable to the here analysed combined trainings. The reported effectiveness of this category is almost equivalent with an effect size of 0.42, based on eight studies. However, the authors failed to report

[28] They distinguished the following components: 1. home visiting component, 2. parent training/child management and/or educational strategies component, 3. preschool programme component, 4. family/parenting support and education, guidance, case management, and referrals to other agencies component, and 5. centre-based childcare/developmental day care component (Manning et al., 2010).

whether this effect differs significantly from zero. It can be concluded that including both children and their parents into an intervention holds great potential for improving the educational development of the children. On a theoretical basis this is underpinned by developmental system theories (Bronfenbrenner, 1979, 1989; Lerner & Castellino, 2002) and findings from research of associated interventions (Nelson, Laurendeau, Chamberland, & Peirson, 2001; Yoshikawa, Aber, & Beardslee, 2012; Zigler, Taussig, & Black, 1992). However, few is known about the circumstances when realising such programmes due to the low number of combined intervention studies. It is imaginable that these programmes are harder to implement since they target multiple parties which requires more personal, more organisation, and more money. Targeting multiple parties also demands more resources from all those involved in the intervention process. Hence, in order to further investigate the conditions of successfully implementing combined interventions, further studies with a high quality design are required.

4.1.4 Long-Term Effectiveness of Preventive Interventions

One challenge for intervention researchers is to achieve sustainable results. Here, the effect on child outcomes across all training types stayed relatively constant up to 12 months after the intervention ended and significantly decreased after that. The effectiveness of child trainings on child outcomes, in contrast, had a peak regarding mid-term effects ($d = 0.57$), and then stayed relatively stable even beyond 12 months post intervention: The treated group still had a 12% significantly higher success rate ($d = 0.24$) than the untreated group. Nelson, Westhues, and MacLeod (2003) found a long-term effect on cognitive outcomes of similar size ($d = 0.30$) when synthesising preschool prevention programmes and judged them as "quite impressive" (p. 22), considering the long time period between the age of the children at programme conduct and at follow-up measurement. However, their longest follow-up was to eighth grade. In the thesis at hand, some follow-ups were assessed more than 10 years after the intervention ended. Consequently, the measurement of the outcomes had to be adjusted to the growing age of the children. For instance, the sample of the *Carolina Abecedarian Project* was about 6 years of age when the intervention ended and IQ was assessed by the Wechsler Intelligence Scale then, and up until the age of 21. At age 21

data regarding years of education and number of high school graduates were additionally collected and at age 30, outcomes such as job prestige and mean income were added. If applying the 12% success rate to income and assuming an income of 2000$ per month for the control group at age 30, the intervention group would receive 240$ more which would be a substantial amount.

The effect of parent trainings on child outcomes was descriptively smaller than child training effectiveness but stable until 12 months after the intervention ended (success rate of 13%). However, beyond 12 months no effect could be found anymore. The same picture could be found regarding parental outcomes across all trainings. For school outcomes, however, already three months after the intervention ended, no effect could be observed anymore. This decrease in intervention effect over time is consistent with other research syntheses on compensatory education intervention (e.g., Barnett, 1992, 2008; Camilli et al., 2010; Chambers, Cheung, Slavin, Smith, & Laurenzano, 2010; Kim & Quinn, 2013).

4.2 Moderating Influences of the Mean Effectiveness at Post Time-Point

Next on the agenda are the moderations of the average effect which will be separated in three sections. Initially, the moderation effects for parental outcomes when implementing parent trainings will be discussed, followed by two sections where the moderating influence for child and parent trainings will be contrasted. This will be done by first discussing the moderations regarding educational child development and then going more into details by looking more differentiated at basic and school development improvements.

4.2.1 Parent Training Effectiveness Regarding Parental Outcomes

Only very few variables actually moderate the effectiveness of parent trainings on parental outcomes: None of the methodological moderators influenced the effect size, but one sample moderator (i.e., proportion of dominant ethnicity), and one intervention moderator did (i.e., implementation problems).

Proportion of Dominant Ethnicity. With higher ethnical homogeneity, the effectiveness of parent trainings increased (β = 0.51). It could be assumed that ethnically homogeneous groups share similar experiences in child care and, hence, have more understanding for each other. That, in turn, could create a particularly constructive and open-minded atmosphere, boosting the effect of the intervention. This hypothesis was supported by the finding that homogeneous samples also had more contact between the intervention and the control group which, surprisingly, seems to have boosted the intervention group (instead of equalising the control group to the intervention group, as one would expect). The influence of exchange between the two groups will be further discussed in section 4.2.2, paragraph ***Contact Between the Experimental Groups***. A corresponding tendency could be observed among child outcomes, as well. Regarding child training, however, ethnically homogeneous samples had similar effects on child development as heterogeneous ones. This result is not surprising considering the everyday experiences of teachers in teaching ethnically diverse children and adolescents and the relevant skills evolving from that. Furthermore, it is highly imaginable that inter-ethnical contact and exchange is more likely at school than within the home environment. Thus, both experienced teachers as well as the exchange between students, might have reduced the predictive value of the variable proportion of dominant ethnicity for child development outcomes.

The type of dominant ethnicity (e.g., African-American versus Hispanics) did not significantly moderate the effectiveness for any of the training types and outcome categories. Interestingly, this finding stands in contrast to comparable meta-analysis. Nelson et al. (2003), for instance, found higher effect sizes among African-Americans as opposed to other ethnicities. However, they included children from various socio-economic backgrounds and did not control for socio-economic status when analysing the effect of ethnicity. As Lee, Brooks-Gunn, and Schnur (1988) puts it, "African-American children were more likely to be big gainers than White children because they were relatively more disadvantaged demographically..." (p. 163). Another recent meta-analysis conducted by Darrow (2009) showed that children with minority status (defined as samples which include predominantly African-American, Hispanic, or Asian/Pacific Islander) profit more from a curriculum intervention than White children. It should be considered, though, that the analysis was restricted to vocabu-

lary outcomes and there were only four comparisons in the category *White*. Thus, this conclusion is based on narrow outcome measures and little evidence. In contrast to Nelson et al. (2003) and Darrow (2009), but accordingly with the results in this thesis, Slavin, Lake, and Groff (2009) and Slavin, Lake, Cheung, and Davis (2009) did not find ethnic background to be a moderating variable for their mean effect.

Implementation Problems. A rather counter-intuitive result was that parent trainings were 14% more effective if problems occurred during programme conduct than if not. This was contrary to the expected effect direction (see section 1.6). Further investigations did reveal that interventions with implementation problems were also ethnically more homogeneous. It is conceivable that ethnically more homogeneous groups might create a stronger in-group feeling and by that a more reserved attitude towards the intervention content, eventually, even be the cause for problems regarding intervention implementation. At first sight, this would be expected to result in a decrease in intervention effectiveness. However, it is possibly that a stronger in-group feeling puts the participants in a position to feel more self-confident and to communicate their reservations concerning the intervention with the trainer. Assuming the trainer constructively seized this opportunity to discuss the theoretically and practically well derived intervention contents, justifying the techniques, and, by that, convincing the parents for good. This could explain the higher effectiveness of the programme, especially if intervention problems occurred but it is also possible that a reporting bias lies behind this finding: Primarily those authors might report problems during programme realisation who conduct high quality research, in the sense of self-critique and transparency. Thus, those studies which reported implementation problems were still better implemented and thus more effective than those who didn't mention problems during intervention implementation. Another imaginable explanation lies in how parent trainings were realised. Usually they were relatively diverse regarding applied methods (on average three to four different ones), with personal contact and relationship as main method by conducting home visitations, and they were of mid- to long-term intervention length. Maybe, in this multi-method, personal, and long-term intervention context, problems during programme conduct lured constructive adaptation processes since the parents are personally more involved and show more initiative, and so does the trainer. This could have boosted the

identification with and motivation for the intervention and, in turn, led to a higher effectiveness.

4.2.2 Child and Parent Training Effectiveness Regarding Educational Child Development

Regarding child development, several interesting influencing variables could be identified for child and parent trainings. Those will be discussed accordingly to the structure of the results section: starting with methodological moderators (i.e., baseline equivalence of intervention and control group, contact between the experimental groups), followed by sample moderators (i.e., children's age, proportion of low-SES children in the sample), and finalising with intervention moderators (i.e., length, density, social deprivation of the area, structuredness and manualisation, implementation problems, as well as training, supervision, and observation).

Methodological Moderators

Baseline Equivalence of IG and CG. Group equivalence showed to be a significant moderator among child trainings in the direction that if intervention and control group differed at pretest, the average effect size was twice as high as if the groups were equivalent. This result is rather difficult to interpret. If this finding was to be trusted, the question would arise how baseline differences could have possibly resulted in higher effects than baseline equivalence. One option could have been that the intervention group had lower scores than the control group and, hence, more potential for improvement. Unfortunately, no variable was assessed here which could answer this question. And even if this information was available, the questions would remain *why* the intervention group started off lower despite the randomised allocation of the participants to the two groups. One possible explanation could be some sort of bias during the randomisation process or the pre-test assessment. What should be mentioned, though, is that the findings also showed that it was more than twice as likely to find individually allocated samples among comparisons with equivalent groups, compared to non-equivalent groups. Further, the effect vanished if group equivalence was analysed separately for individual- versus cluster-randomised comparisons. Hence, it can be concluded that the moderating effect of

baseline seems to be just an artefact due to covariations with other variables.

Contact Between the Experimental Groups. Another moderator of methodological character, this time for parent training effectiveness on child outcomes, was the amount of contact between the intervention and control group during programme implementation. Surprisingly, the findings suggest that interacting groups profited twice as much from the programme as non-interacting groups. Eventually, this could be connected to the consolidating and memorising effect of repeating knowledge and sharing experiences (Baddeley, Eysenck, & Anderson, 2009). The interaction might have increased the consolidation of what was learned and, thus, boosting the intervention effect. Contact seems to play a minor role for child trainings. Maybe, if intervention and control group children, on average 5 years and 11 months of age, interact, the subject is not so much the experiences and knowledge gained from the intervention, but rather social and emotional exchange. What further needs to be considered, though, is the indirect effect of other variables with which contact might be confounded. For instance, socially deprived area was associated with contact in the sense that among parent trainings which were conducted in socially deprived areas, it was also more likely to find interacting groups. As discussed further below in paragraph ***Social Deprivation of the Area***, children living in deprived environments might have more potential for improvement and, hence, reach higher effects. Further analyses revealed that contact even lost its predictive value if it was included together with deprivation of the area into one meta-regression, whereas deprivation was marginally significant. Thus, it seems that it doesn't make a difference on child outcomes whether intervention and control groups interact or not if controlling for socially deprived area, be it about intervention content or not.

Sample Moderators

Among the sample moderators, two interesting findings were observed which will be discussed in detail here, regarding children's age and the proportion of low-SES children in the sample.

Age. The discussion about the influence of age on intervention effectiveness is controversial and "...not conclusive...", as Burger (2010, p. 159) summarised it (e.g., Bakermans-Kranenburg, van IJzendoorn, Marinus H., & Juffer, 2003; Beelmann &

Schmitt, 2012; Burger, 2010; Hahn et al., 2007; MacLeod & Nelson, 2000; Mortensen & Mastergeorge, 2014; Sweet & Appelbaum, 2004; S. J. Wilson, Lipsey, & Derzon, 2003).

Here, age has been found to influence child training effectiveness in the way that an increase in age resulted in lower child outcome effect sizes (β = -0.19). As assumed already by Lösel and Beelmann (2003), those diverse findings could be due to the association of age with characteristics of the intervention. And indeed, when controlling for structuredness (among structured programmes the mean age was lower as opposed to unstructured ones), age did not influence child development anymore. Among parent trainings, no substantial association of age and characteristics of the intervention could be observed and age had no moderating influence. It is interesting to observe that some of the most recent meta-analyses in this field which included children of different ages, didn't investigate its moderating effect or at least didn't publish the relevant results (Darrow, 2009; Kim & Quinn, 2013; Manning et al., 2010).

Proportion of Children with Low-SES in the Sample. The same argumentation as for age can be applied to the finding that the amount of children from low-SES within the sample influenced child development improvements when conducting child trainings: Higher percentages resulted in higher effectiveness. The moderation disappeared, though, if including structuredness as additional predictor. Hence, the moderation of proportion of low-SES was mediated by structuredness.

Thus, neither age nor the proportion of low-SES children in a sample moderated child or parent training effectiveness on child outcomes when controlling for structuredness of the intervention.

Intervention Moderators

Intensity. As with age, the findings regarding the influence of intervention intensity are of controversial character, as well. Some authors found intensity to be of positive influence on the average effect (Manning et al., 2010; Nelson et al., 2003; Yoshikawa et al., 2012), others observed no association or an unclear picture (e.g., Sweet & Appelbaum, 2004), and some even reported that with increasing intensity the effectiveness declined (e.g., Bakermans-Kranenburg et al., 2003). This can partly be attributed to the diverse operationalisations of intensity, for instance, by assessing the number of sessions, the total duration of the programme, the frequency, or a com-

bination of those. Here, it was distinguished between length and density. Those two variables were not related to each other.

Length. Regarding the length of child as well as parent trainings, the findings are in line with those published by Bakermans-Kranenburg et al. (2003): Shorter is better. With longer intervention time the effectiveness decreased and stagnated at a level of d = 0.24 and 0.11, respectively. One explanation could be fatigue, since a longer duration could mean more repetition and routine which leads to boredom and, hence, manifests itself in lower effectiveness. Another explanation could be the confoundation of length with attrition, as suggested by, for instance, Beelmann and Schmitt (2012). Indeed, the findings show that longer programmes also had higher attrition rates among child trainings as well as parent trainings. Hereby, the mean attrition was comparable for the intervention and the control group. Nevertheless, the attrition rate in the control groups for both training types had a descriptively higher predictive value regarding child outcomes than the drop-out in the intervention groups. Thus, it could be argued that with increasing length of the intervention the participants' motivation in the control group might have decreased which resulted in higher attrition rates of particularly unmotivated participants, leaving primarily the motivated ones behind. This led to smaller effect sizes because, as Bakermans-Kranenburg et al. (2003) express it, "...it would be more difficult for the experimental group to outperform the control group" (p. 205).

Density. In contrast to length, the intervention effectiveness among parent trainings increased the more dense the intervention sessions were clocked. It stagnated at 14 sessions per month with an effect of d = 0.50. However, when controlling for social deprivation of the area, density had no significant moderating effect anymore. The influence on child outcomes of whether an intervention was conducted in a socially deprived area or not will be discussed further below in paragraph ***Social Deprivation of the Area***. Among child trainings, a trend could be detected toward a reversed U-shaped relation: The effectiveness increased up to a density of 13 sessions per month, reaching a mean effect size of 0.50 and then decreased again. However, this finding should be interpreted with caution, since it was only a trend. Furthermore, a covariation of density with site of study was detected. Among interventions with high density it was more likely to find multi-site studies, whereas single-site studies were

typically of medium density. As already discussed in section 4.1.1, the lower average effect sizes which multi-site studies achieve, were connected to intervention characteristics, more precisely to manulisation as well as social deprivation of the area.

Implementation Fidelity. As hypothesised in section 1.6, it was expected that well implemented interventions should result in higher effect sizes than if implementation fidelity is not given because they were theoretically better derived and, thus, more likely to be implemented as designed. The findings regarding this hypothesis will be discussed in three parts: structuredness and manualisation, implementation problems, as well as training, supervision, and observation.

Structuredness and Manualisation. Regarding structuredness, the findings confirm the hypothesis: Structured child as well as parent trainings were 3.7 and 3 times more effective, respectively, than unstructured programmes. It is important to emphasize here that unstructured interventions had, statistically, no effect at all on child outcomes, neither for child nor for parent trainings. Hence, in order to achieve improvement in children's educational development it seem that a basic structure is necessary or more precisely, the topics should be roughly predetermined and the number as well as the length of the sessions should be scheduled.

In contrast to structuredness, manualisation only moderated the effectiveness of parent trainings in the expected direction: Manualised programmes were 5.7 times as effective as unmanualised ones. No moderation by manualisation could be found for child trainings. However, there was a stronger association between manualisation and structuredness among parent trainings as opposed to child trainings. Precisely, it was much more likely to find unmanualised but structured interventions among child trainings than among parent trainings. One explanation for this could be that a reporting bias was involved. In school contexts, were child trainings were usually implemented, curricula or manuals are omnipresent as a tool to organise the learning content. Hence, its application might have seemed self-evident and so the authors did not mention applying a curriculum or manual in their study even if it was used. Consequently, many manualised interventions might have been coded as unmanualised which increased the mean effect among the latter and, hence, reached a similar effect level as manualised programmes. If controlling for structuredness, manualisation lost its predictive value and structuredness alone moderated the child outcome

effectiveness, indicating it to be a more relevant moderator for child training effectiveness. Among parent trainings, in contrast, manualisation crystallised as predictor for effect size variations. As already mentioned in section 4.2.1, parent trainings were of rather personal character. Thus, the intervention effect might depend stronger on the relationship between parents and the trainer. Hence, a rough framework, in form of a manual, boosted the intervention effect. However, additional detailed regulations regarding when to do what (i.e., structuring the intervention) did not further increase the effectiveness.

Implementation Problems. In accordance with the findings regarding structuredness and manualisation, if interventions were not implemented as planned, child trainings were only half as effective as if no problems occurred during conduct. Regarding parent trainings, no significant moderation could be found regarding child outcomes but for parental outcomes a moderation could be observed in the reversed direction which was already discussed in section 4.2.1, paragraph ***Implementation Problems***.

Training, Supervision, and Observation. Regarding training, supervising, and observing the executor of the intervention it was hypothesised to find a positive influence on programme effectiveness, compared to not applying those techniques. The findings indicate, though, that solely training and not supervision or observation, and only if a parent training was implemented, was superior to not training the executor (d = 0.32 versus 0.12, respectively). Supervision and observation did not moderate the child outcome effect among parent trainings. In contrast to that, child trainings were by 7 and 9% *less* effective if the executor was supervised and observed, respectively, compared to if no supervision or observation was conducted. Hence, observation and supervision seem to have been obstructive for child training effectiveness. Descriptively, similar findings could be revealed regarding training of the executor (a 6% higher success rate for untrained executors than for trained ones). However, the moderation did not reach significance.

Since most of the child trainings were realised by teachers in the school context, a possible explanation for those findings could be that teachers, as well-trained and experienced knowledge mediators, were disturbed in their normal routine by observation and supervision, causing insecurity and, by that, decreasing the intervention

effect. Furthermore, when keeping in mind that the school curriculum is already a relatively highly structured and organised setting which demands a lot of resources from teachers, it is imaginable that changing, adding or replacing all or parts of their daily routine might lead to higher stress levels due to necessary adaptation and learning processes. They might simply not have enough resources available for being supervised and observed on top of conducting an intervention. This theory of limited resources is supported by the findings that supervision and observation lost their significant negative moderation if only interventions were analysed which did not take place inside the educational setting or which were not conducted by educational staff. Moreover, when only analysing programmes conducted over a long time period (duration of more than 12 months), the negative effects diminished, as well, suggesting that routine was established taking up less of the resources but still not resulting in positive effects. In contrast to child trainings, parent trainings were typically implemented via one-on-one interactions and, hence, rather worked by establishing a personal relationship. Training of the trainer obviously is a valid and necessary technique for developing the competencies needed to implement the programme. Interventions with untrained executors had statistically no effect. However, it seems unnecessary or even contra-productive to supervise or observe the trainer during programme conduct.

To the author's knowledge, there exists no recent meta-analysis which systematically investigated the influence of implementation variables on programmes aiming to improve the educational development of children and adolescents, especially not for children of low-SES. The closest scientific topic where such analyses can be found, target the children's socio-emotional development by programmes that, for instance, target bullying, drug consumption, or aggressive behaviour (Derzon, Sale, Springer, & Brounstein, 2005; DuBois, Holloway, Valentine, & Cooper, 2002; J. D. Smith, Schneider, Smith, & Ananiadou, 2004; Tobler, 1986; S. J. Wilson et al., 2003). The meta-analysis conducted by DuBois et al. (2002) synthesised almost 500 of such studies and did assess multiple types of outcomes, among them academic/educational outcomes. However, the results regarding fidelity were just reported over all outcomes and revealed that the "...mean effect sizes are at least two to three times higher when programmes are carefully implemented and free from serious implementation problems

than when these circumstances are not present" (p. 340). Comparable and even partially higher effects regarding implementation fidelity in this thesis could be observed for child trainings on child outcomes concerning structuredness (3.7 times higher if programmes were structured) and implementation problems (2.2 times higher if no problems occurred). For parent trainings on child outcomes that was the case for manualisation (5.7 times higher if a manual was provided) and training (2.7 times higher if executor was trained).

In a nutshell, child training effectiveness was higher if the intervention was structured and if no problems in implementing it were reported. Regarding the moderators supervision and observation of the executor during programme conduct, a negative effect on child outcomes could be observed. Among parent trainings, manualisation did predict child outcome effectiveness in a positive way and training of the executor was indeed found to be essential for achieving child outcome improvements. But neither supervision nor observation had substantial influence here.

Social Deprivation of the Area. Another interesting result is that the intervention effectiveness among parent trainings was 2.5 times higher if the programme had been conducted in a socially deprived area as opposed to non-deprived areas. The area where the intervention was conducted can be expected to be located close to the participants home due to economic reasons, such as transportation. Thus it can be assumed that if the intervention was conducted in a socially deprived area, it was very likely that the participants lived in the corresponding area, as well. As described in detail in section 1.3.2, socially deprived areas have higher violent and household crime rates and, thus, the children growing up there are more often exposed to violence (Dodge et al., 1994; Federman et al., 1996; Garbarino, 1999). Due to a lack of resources, the access to high-quality services (e.g., schools, child care, hospitals) and informal social support is lower (Lee & Burkam, 2002; McLoyd, 1998; Phillips et al., 1994). Thus, children and adolescents living in areas with social deprivation do have fewer or qualitatively lower opportunities inside and outside their home environment compared to those living in non-deprived areas. Before the beginning of the intervention those children might have already had lower cognitive or achievement scores and by that more potential for benefiting from the intervention which resulted in greater educational improvement. It is also reasonable to presume that children

in more deprived living circumstances soak up learning opportunities, such as the conducted interventions, like a sponge, leading to higher intervention effects. Kim and Quinn (2013) argued in a similar way. They explained their observed higher average effect among low-income children as opposed to middle-income children by the absence of learning opportunities in their home environment (e.g., fewer number of books available, parents spending less quality time with children). According to Bronfenbrenner's ecological systems theory, neighbourhood is, like home environment, a developmental context for children only assigned to the exosystem instead of the microsystem (Bronfenbrenner, 1979, 1989).

4.2.3 Child and Parent Training Effectiveness Regarding Basic and School Development

The effectiveness of child and parent trainings on basic and school development will be discussed in this section. Hereby, the average effects will be looked at first, followed by a more differentiated evaluation regarding selected moderators for child and parent trainings, separately.

Correspondingly to the findings across all child outcomes (see section 4.1.3), approaching children directly yielded higher intervention effects for basic (d = 0.42) as well as school development (d = 0.32), as opposed to when indirectly addressing the children through their parents (d = 0.32 and 0.23 for basic and school development, respectively). Hence, basic and school development improved about 5% more if a child training was implemented as opposed to a parent training. Furthermore, descriptively higher basic development improvements as opposed to school development could be observed for child and parent trainings. A comparison of these findings to those of similar meta-analyses is rather limited because the constructs are measured differently. Camilli et al. (2010) as well as Manning et al. (2010), for instance, did distinguish between several outcome constructs. Camilli et al. (2010) built a *cognitive*, *school*, and *social domain*, as they referred to it, by considering contextual similarities as well as the mean differences between the original and the built effect sizes. Since Camilli et al. focused on centre-based early education interventions, the effect regarding school domain can be considered roughly equivalent to the here derived school development effect size when conducting child trainings. The mean effect size

reported by Camilli et al., was not even half the size ($d = 0.15$) as the one found in the meta-analysis at hand, though. As already mentioned in section 4.1.2, Camilli et al. did not restrict the sample to children with low socio-economic background. Hence, the lower effect size can be explained by the higher improvement potential of children living in disadvantaged environments as already discussed in section 4.2.2, paragraph ***Social Deprivation of the Area***. Furthermore, Camilli et al. assigned achievement outcomes to the cognitive domain which could have also influenced the average effect on the school domain. Even though this makes the cognitive domain limitedly comparable to the here derived basic development construct, Camilli et al. also found higher effects on the cognitive ($d = 0.23$), compared to the school domain as it was found in this thesis, as well.

Manning et al. (2010) created seven outcome domains by reviewing the psychometric literature. Among the domains are, as they referred to them, *cognitive development* and *educational success*. They also assigned non-cognitive outcomes, such as school grades or achievement measures, to the cognitive domain. Thus, again, the constructs are only comparable to a limited extent. Moreover, the constructs were not analysed separately for child and parent trainings. The effects they reported on the cognitive and educational success domains were relatively high ($d = 0.34$ and 0.53, respectively, compared to $d = 0.37$ and 0.28 for the meta-analysis at hand across child and parent trainings), considering that the sample consisted of disadvantaged, but not necessarily of children with low socio-economic background. The medium effect size regarding educational success is especially surprising. However, as already argued in section 4.1.2, only studies which measured a follow-up during adolescence were considered for inclusion in the meta-analysis of Manning et al. which might have lead to the inclusion of especially effective programmes and, thus, explain the strong effect.

Child Training Moderators. Among child trainings, moderator analyses revealed that structuredness, supervision, observation, and implementation problems moderated school development effectiveness correspondingly to the findings regarding overall child outcomes for child trainings (see section 4.2.2). Structured interventions were more than 4 times as effective as unstructured ones. Unsupervised and unobserved programmes were 1.4 and 1.6 times as effective as supervised and ob-

served ones. Implementation problems decreased the effect to half, compared to if no problems during programme conduct occurred. It should be mentioned, though, that the finding concerning supervision should be understood as a trend, since it only marginally reached significance. In contrast to the moderation of school development effectiveness, none of the investigated variables had a substantial moderating effect when analysing basic development outcomes. The only exception could be observed for implementation problems, where a trend was revealed of a 2.6 times higher effectiveness if the intervention was well-implemented. This finding did not reach significance, though, due to the low number of interventions reporting implementation problems ($k = 4$). Hence, it was confirmed for basic as well as school development that if the intervention was implemented as planned, child training effectiveness was higher as if problems during conduct occurred.

In order to understand the contrasting findings for basic development and school development regarding structuredness, supervision, and observation, the participants' age could serve as an explanation. If analysing the influence on basic development, the mean age was 4 years and 10 months, compared to 7 years if school development was investigated. When looking at basic development, it was further more likely that methods which require one's own initiative, such as free/guided play or performance (i.e., playing theatre, drawing, painting, reading together), were applied as opposed to school development. According to Erikson's Stages of Psychosocial Development (Erikson, 1963/1995), children between 4-5 years of age pass through the phase of *initiative versus guilt*. At this stage, the children's independence and courage increases. They more often take initiative and engage in planned activities. This might demand the teachers to provide the children with diverse and open opportunities and stimulation, rather than applying ordered and organised techniques. This could explain why structuredness did not significantly moderate basic development effectiveness.

Regarding the negative influences of supervision and observation on child outcomes for child trainings, teachers' resources were discussed as possible explanation (see section 4.2.2, paragraph ***Training, Supervision, and Observation***). In contrast to school environment, preschool environment leaves teachers with more resources and higher flexibility regarding design possibilities due to, for instance, less restric-

tive and simpler curricular requirements, with regards to content. This might explain why supervision and observation did have no negative influence on basic development improvements.

Parent Training Moderators. Manualisation proved to be essential not only for child outcome improvements, but also for achieving basic and school development improvements: If no manual was applied, parent trainings had no significant average effect, whereas a small and substantial effect could be found if a manual was provided (d = 0.39 and 0.34, respectively).

The same can be concluded for training the trainer regarding basic as well as school development: Only if a training was implemented, a small but substantial effect could be achieved (d = 0.31 and 0.35, respectively). For basic development, however, this difference was not significant. Those contrasting findings for child and parent trainings, might be explainable by the differing average age of the two training types when analysing basic or school development. The mean sample age was 1 year and 6 months if basic development was investigated and 3 years and 6 months if school development was analysed. According to Erikson, parenting revolves around meeting the child's basic needs when children are between 0 and 2 years of age and, by that, creating a trustworthiness. He called this first stage *basic trust versus basic mistrust* (Erikson, 1963/1995). This requires personal characteristics, attitudes, and behaviours of the parents, such as warmth, protection, dependable affection, and regularity. It is reasonable to assume that for attending those needs and influencing them by an intervention, an empathic relationship between trainer and parents is essential. Establishing that, is hard to teach a trainer which is why training might have had no influence on basic development. Children at the age of 2-4 years, though, are in the stage of *autonomy versus shame and doubt*. Their motor skills, speech, and understanding develop rapidly and so does their need for exploring the environment. By doing that, they experience their growing sense of autonomy but are, at the same time, still very dependent on their parents. It is the parents' task to encourage the children but at the same time not to demand too much of them. In this context, Erikson spoke of a surplus of energy that children have at that age which "...enables them to forget failures quickly and to approach new activities with undiminished enthusiasm" (Hook, 2009, p. 291). Parents can promote this enthusiasm, assuming they under-

stand their children's capabilities and are able to constructively communicate with them. Providing parents with knowledge or techniques how, for instance, to communicate with the child or how to promote their child's educational development goes beyond empathy and needs to be trained in order to achieve an intervention effect. Hence, an untrained executor does not put parents in the position to enhance their child's school development, as the findings showed.

Furthermore, parent trainings were especially effective regarding basic development if the intervention was conduced in a socially deprived area, reaching a substantial medium effect ($d = 0.55$). In fact, interventions implemented in socially deprived areas were 3.4 times as effective as those conducted in non-deprived areas. A corresponding trend could be observed regarding school development but the moderation did not reach significance. One explanation for the ceasing moderation and the descriptive effectiveness decline for parent training conducted in deprived areas from 0.55 for basic development to 0.33 for school development, could be the foundation of numerous preschool programmes in the United States since the 1960s, especially for disadvantaged children. The two most prominent examples are Head Start and the High/Scope Perry Preschool Project. Regarding Head Start, the enrolment rate in the full-time programme in the 1970s was about 200 000 and increased with the years until it almost reached 1 million participants in 2003, where it stabilised (Early Childhood Learning and Knowledge Center, 2016). Hence, it is likely that 3 $^1/_2$-year-old children, attended (some form of) preschool, particularly those of low-SES who are more likely to live in socially deprived areas. Those children already achieved educational improvements and by that decreased the negative influence of living in deprived areas. Thus, the moderator ceases to influence the intervention effect.

4.3 Recommendations for Research

4.3.1 Expanding Research Regarding Implementation Fidelity

Not all variables expected as benefiting for implementation fidelity did actually moderate the intervention effect in the expected direction. Some even had reversed influences (i.e., supervision and observation for child trainings) but it is scarcely possible to investigate the processes which might lie behind those findings or to compare

them with those of other syntheses works in this field. One reason for this is that the reporting of circumstances regarding an intervention's implementation is a relatively recent development, and so is the investigation of this concept as a moderator in meta-analyses. Furthermore, due to its recency, its operationalisation is still extremely diverse. This can be observed very well in a special issue which Griffin (2010) published, presenting different approaches which were already applied in research with the aim to explore implementation. Among those were, for instance, the exploration of patterns concerning teacher participation, dosage, adherence, quality of delivery, videotaping as observational method and for feedback, fidelity, generalisation, and child engagement. Griffin strikingly concluded that, "[quantitative] science cannot study what it cannot measure accurately and cannot measure what it does not define" (p. 342). It can be stated that the importance of implementation characteristics is recognised by research today. Durlak and DuPre (2008), for instance, fittingly stated that, "...the assessment of implementation is an absolute necessity in programme evaluations" and that intervention research which does not collect data regarding this topic should be considered "...flawed and incomplete" (p. 340), since the findings cannot be interpreted reliably. Theoretically, this problem has been approached by scientists resulting in, for instance, the inclusion of reporting implementation problems into the Reporting Standards for Research in Psychology (JARS Group, 2008). In 2013, Simpson et al. even published a tool for evaluating research implementation challenges (TECH). Now those theoretical standards need to be realised in research practice and in article reporting and, by that, allowing to further investigate and enhance research on implementation characteristics.

Variables regarding the implementation of the intervention are further of potential interest for unravelling the association between sample size and effect size, as discussed in section 4.1.1.

4.3.2 Defining and Operationalising Outcomes

The formation of the outcome constructs in this meta-analysis was based on theories of the renowned psychologists Jean Piaget and Lev Vygotsky (see section 1.4.3). In contrast to this theoretical derivation, the use of the term *cognitive development* in intervention research is often diluted. One reason for that is the rather subjective and,

thus, questionable approach of classifying the outcomes in many meta-analyses (see section 4.2.3). The only common approach is, though, that cognitive development in the context of intervention research typically includes IQ measures. Apart from that, it has to be judged as a smorgasbord of different measures, sometimes including achievement, in other cases adding school grades or both (Camilli et al., 2010; Manning et al., 2010; Nelson et al., 2003). This arbitrariness in differentiating outcomes is a threat to the conduct of high quality research. This applies in particular to meta-analyses: Without a common outcome taxonomy, a comparison of the findings between different meta-analyses on similar topics, is hardly feasible. The outcome structure should be based on theoretically well thought-through definitions. This meta-analysis suggests one option in how that can be achieved.

4.3.3 Hispanic Sample

As mentioned in section 3.1.2, an unexpectedly high number of primarily Hispanic samples was part of this analysis. Those studies were retrieved by a meta-analysis with similar eligibility criteria (see section 3.1.1) which was conducted parallel, with the exception that it focused on samples with migration background and not with low socio-economic background. However, some of those met the eligibility criterion for low-SES and were, thus, included in the study pool. This was the case for 10 of the 109 publications representing 9% of the study pool, which is a considerable amount. Those studies could not have been found by using search terms for children of low socio-economic background. This ethnic group is underrepresented in meta-analyses which investigate disadvantaged children and adolescents (e.g., Darrow, 2009; Manning et al., 2010). It is therefore important to explicitly consider Hispanic samples in future synthesis research when looking for participants of low-SES by adapting the search strategy.

4.3.4 Comprehensive Data Analysis

The study pool of 109 publications allowed the conduct of comprehensive and detailed investigations, especially regarding interrelations between the different moderators. The presented findings show the importance of such analyses for unbiased

and reliable conclusions concerning intervention effectiveness. The finding, for instance, that the older the children are the fewer the interventions effect them was falsified, once structuredness was added as an additional predictor. This depth is certainly only possible when writing a dissertation or comparable theses. However, even when publishing journal articles, covariation between the tested moderators should be considered. It is reasonable to assume, that many controversial results might be explainable if confoundations with other variables would have been tested and reported. Consequently, future meta-analyses should focus more on testing and reporting on covariations between the assessed study characteristics.

It needs to be considered, though, that multiple testing increases the probability of finding false positive results (i.e., Type I error). The risk can be reduced by reasonably selecting moderators, for instance. This issue is further discussed in section 4.5.6.

4.3.5 Research Gaps

Regarding child and parent trainings, the number of studies was sufficient to conduct comprehensive analyses. This was not the case for teacher trainings and combined interventions. Especially, combined interventions have promising, almost medium effects. However, more such trainings need to be conducted and evaluated in order to understand optimal programme conditions.

4.4 Recommendations for Practice

Recommendations for practise will be given regarding the implementation of parent and child trainings targeting to improve basic and school development. Beyond that, possible structural and sustainable changes will be suggested.

4.4.1 Preventive Child Training Implementation

A number of factors emerged in this thesis which should be considered when implementing child trainings with the aim to improve the children's educational development. First of all, educational improvements can be achieved regardless of the children's age and the proportion of children with low-SES among the participants. Furthermore, neither the frequency of the sessions, nor the length of the programme

was found to be of practical influence (when controlling for other factors). In contrast to that, it is essential to structure the training content if targeting child development. The programme effect is 3.7 times higher for structured than for unstructured programmes. A supervision or observation of the trainer should not be conducted - or at least not in the way they were implemented so far. Those measures seem to disturb the intervention rather than improve it. Problems during intervention conduct should be prevented: Interventions which can only be partly implemented as planned don't reach their full potential.

When aiming at enhancing children's school development, all the described recommendations can be transferred. Regarding children's basic development, however, none of the above mentioned factors need to be considered, except for implementation problems. For achieving improvements, it is important to ensure that the intervention is implemented as planned.

4.4.2 Preventive Parent Training Implementation

As already stated in the context of child trainings in section 4.4.1, parent trainings enhance child development independently of the children's age, the proportion of children with low-SES among the participants, the length, or the density of the intervention. An important factor to consider, though, is the social deprivation of the area. If intending to achieve higher intervention effects, not generally children and adolescents from low socio-economic background should be targeted, but precisely those who live in deprived neighbourhoods. Furthermore, for achieving developmental changes in children by parent trainings, providing a manual and conducting a training with the trainer are essential.

If specifically aiming at enhancing children's basic development, the application of a manual continues to be fundamental but a training of the executor can be disregarded. The deprivation of the neighbourhood should also be considered. If parent trainings are intended to improve children's school development, the application of a manual and training the trainer are essential. The circumstances of the neighbourhood the children live in, can be disregarded in this context.

4.4.3 Changing Structural Disadvantage

As the findings show, preventive interventions for low-SES children and adolescents just compensate the educational development delay to a certain degree (see section 4.1.2). What can be further be concluded from other studies is that especially pre-kindergarten programmes for children from low socio-economic background seem to be beneficial in reaching national norms in school readiness (see section 4.1.2). Beyond pre-kindergarten the compensating value is rather disillusioning and only promising for isolated areas, namely pre-reading skills. Thus, the application of preventive interventions for closing the educational performance gap between children of low-SES and the national norm remains questionable.

Realistically speaking, nothing more can be expected since the programmes only provide support in usually isolated developmental areas for a limited time. After the intervention ended, the children are left alone, again, in living circumstances which are less advantageous since they are more likely to live in socially deprived areas or in hardly stimulating home environments, and attend public schools which receive less (social) support (see section 1.3.2). Thus, this structural and persistent disadvantages cannot be overpowered by selective and timely limited interventions.

If policy makers truly intended to close the developmental gap between poor and non-poor children for good, the approach would need to be sustainable and comprehensive, involving structural and political changes, as well as attitude changes regarding the teacher job as a respected and aspired job option. The findings show that by implementing teacher trainings, no substantial child development improvements can be achieved. However, as already discussed in section 4.1.3, teachers have a high sphere of influence and, hence, the potential for enhancing children's basic and school development.

As an inspiration on how that could be realised, Finland could serve as a model. The country ranges in the top positions of international school rankings since their introduction (e.g., PISA; OECD, 2013). In the 1960s and 1970s, Finland underwent major structural transformations including also the educational system (Antikainen & Luukkainen, 2008). Today the country is amongst the highest ranked countries on the Education Index (United Nations, 2014) and has taken worldwide recognition for its high-quality education and equal educational chances for all children (Lie,

Linnakylä, & Roe, 2003). In 2014, Finland belonged to the countries with the lowest rate of children living below the poverty line in Europe[29] (Eurostat, 2016a) and also among the OECD member states in 2013[30] (OECD, 2016b). The same applies to inequality measures, such as the *Gini coefficient* where Finland was at 0.26[31] in 2013 which can be interpreted as relatively even distribution of the income in the national population. Another inequality statistic is the *Top 10% vs bottom 10%* measure which was with 6% the fourth lowest for Finland in 2013. This number signifies how many times the average bottom 10% income has to be multiplied to reach the top 10% average income of the population. This success cannot only be due to higher salaries or the economic situation in form of gross domestic product per capita of the country, since Finland is close to the OECD average regarding those indexes (OECD, 2013, 2016a).

However, Finland differs strongly from most of the other OECD countries regarding their policy on education. Their education system can be regarded as highly equitable because it is free of tuition and taught comprehensively until the children are 16 years of age. School meals are subsidised and full-time schools are the standard (Antikainen & Luukkainen, 2008). Compared to the OECD average, the schools are largely free from external requirements. Decisions concerning the school are made locally, making the principals the "...pedagogical leaders of their schools" (OECD, 2013, p. 12). Moreover, teacher education is highly selective - only 10% of the applicants for primary teacher studies are accepted. The application process contains, for instance, observations in teaching-like activities as well as interviews in which their aptitude for teaching is the subject of discussion (OECD, 2013). They then "...receive a strong theoretical and practical teacher education and are highly respected [in their profession]" (OECD, 2013, p. 10). Teacher students are trained in enhancing their teaching techniques on a scientific basis and develop pedagogical knowledge them-

[29] 11%. The poverty threshold was set at 60% of the national median equivalised disposable income after social transfers. The considered population was less than 18 years of age.

[30] 5%. The poverty threshold was set at 50% of the national annual median equivalised post-tax and transfer household income. The considered population was less than 17 years of age.

[31] The Gini coefficient can rage from *0* which corresponds to total equality (everybody had the same income) to *1* which corresponds to total inequality (a single person had all the income).

selves. They are further enabled to adapt their teaching depending on, for instance, the students' learning needs or their learning style. The numbers show that Finland's education system can serve as a model for breaking the dependence of students background and their educational performance (OECD, 2013).

In the light of the above, the early segregation in the German school system or the financial dependence of high-quality education in the United States should be regarded highly critical and be reconsidered. Theoretically, the topic seems to have reached policy makers in the European Union: In 2006, most of the member states of the European Union explicated in the National Reports on Strategies for Social Protection and Social Inclusion the "...need to develop a strategic, integrated and long-term approach to preventing and alleviating poverty and social exclusion among children" (European Commission, 2008, p. 10). However, as the numbers regarding the devastatingly contradictory development of wealth and poverty show (see section 1.1), the subject seems to remain a theoretical one for now.

It could be argued that financial restraints hinder the realisation of such a holistic approach. However, Finland only spent 1.5% more of its gross domestic product (GPD) on education in 2013, compared to the European Union average (6.5 versus 5.0%, respectively; European Commission, 2015).

Hence, breaking the dependence of students socio-economic status and their educational performance can be achieved by, to express it with the words of Lie et al. (2003), "...constructing a publicly funded comprehensive school system without selecting, tracking or streaming students during their basic education until the age of 16" (p. 8), and by providing them with highly competent and professionally respected teachers.

4.5 Limitations

4.5.1 Focussing on Educational Outcomes

The dissertation focussed on investigating the intervention effectiveness on educational outcomes. That allowed the conduct of thorough analyses. It would have been interesting, though, to compare the effectiveness on different child outcome domains, as Camilli et al. (2010) or Manning et al. (2010) did. Directly comparing changes in

basic and school development with other relevant changes regarding, for instance, social competence measures, internalising and externalising behaviour problems, or adolescents' substance abuse and sexual behaviour, would have resulted in a more holistic picture regarding preventive intervention effectiveness for low-SES youths.

4.5.2 Focussing on Variables on Comparison Level

The moderator analyses of this thesis focused on variables which were assessed on comparison level. However, it would have also been interesting to investigate variables which were measured on effect size level. It could be presumed, for instance, that the intervention effectiveness differed if the authors used a standardised or an ad hoc measurement. Furthermore, there might be effectiveness differences if the outcome was assessed on learning, behavioural, or result level, or whether the outcome corresponded to the aims or contents of the intervention or not.

4.5.3 Time Discrepancy Between Implementation and Publication

One problem which is rarely discussed throughout intervention research in social science, is the time discrepancy between intervention implementation and the publication date of the corresponding meta-analysis. The problems which lie herein particularly affect synthesis research which doesn't limit the primary studies regarding their publication year (such as the dissertation at hand). Especially interventions realised a long time ago (e.g., Di Lorenzo & Salter, 1967) or interventions with follow-ups of 20 years and more (e.g., Ramey, Bryant, Campbell, Sparling, & Wasik, 1988; Weikart, Deloria, & Lawser, 1970) implemented methods or contents which might no longer be applied today. Constant adaptations of the intervention, on the other hand, does only provide limited information regarding long-term effectiveness. Those contradictory demands need to be considered cautiously when discussing findings in intervention research and drawing conclusions for practise.

4.5.4 Reporting Gaps in Primary Studies

Even though there has been a lot of improvement regarding reporting standards (e.g., JARS Group, 2008), many reporting gaps still exist, even among *classical* study

characteristics. For instance, it was sometimes even difficult to gather all relevant information concerning the randomisation procedure. Furthermore, the intervention process appeared occasionally unclear, concerning the beginning and ending time point, how many sessions in what time span were realised, and how long the conducted sessions were. Regarding the variables which assessed programme implementation, it was assumed that if no information was reported (e.g., regarding training or supervision), the criterion did not apply (e.g., no training or supervision was conducted). This is a conservative assumption without which a substantial amount of studies would have otherwise been rated as missings which would have jeopardised the statistical power. The assumption might, though, have lowered the reliability of the findings. For most variables, however, sufficient descriptive information was available to rate them with relative certainty. Still, it is necessary and essential in some areas, such as implementation fidelity, to work out more specific guidelines for meta-analyses based on common vocabulary.

4.5.5 Generalisability of the Results

Even though no limitations were set regarding publication language and country of origin, almost 90% of all studies were conducted in the United States. This limits the generalisability of the findings. Even though, compensatory early education programmes exist throughout the world (e.g., Effective Provision of Pre-School Education (EPPE) in the United Kingdom, Dutch Cohort Study of Primary Education (PRIMA) in the Netherlands, or Early Childhood Development in rural Vietnam), many publications did not meet the minimum methodological standards of this meta-analysis. It can be argued that lowering those standards would result in a more representative study pool for drawing worldwide conclusions. However, certain methodological standards are important to ensure the internal validity of the results. If internal validity is threatened, the conclusions about the intervention effectiveness, as well as their generalisation are in jeopardy (Cook & Campbell, 1979a).

Regarding this limitation, Manning et al. (2010) point out that their findings on early developmental prevention should be generalised with caution due to the particular circumstances in the United States. They further mentioned that, "...the recent positive evaluations of Sure Start in the United Kingdom and Communities for Chil-

dren in Australia [...] suggest that generalisation to other English-speaking countries may be quite feasible" (p. 516). This conclusion can be regarded transferable to the meta-analysis at hand. The question whether the findings can be generalised to non-English-speaking countries, though, remains.

However, considering the higher effectiveness of interventions conducted in deprived areas, it could be assumed that the intervention effect seems transferable to undeveloped and developing countries because of their higher potential for improvement[32].

4.5.6 Exploratory Analyses

Regarding the analysis of the moderators, many statistical tests were conducted. For instance, each variable was tested for covariation with all the remaining ones, separately for child and parent trainings. This was done in order to find confounding effects. That increased the likelihood of a Type I error, in other words, the probability of incorrectly rejecting the null hypothesis augmented. One possible solution for this problem is a Bonferroni correction via adaptation of the critical p value by dividing it by the number of hypotheses tested (Abdi, 2007). However, this would have resulted in a critical p value of p = .000014[33]. None of the findings would have then reached significance which is unrealistic as the comparison to the findings of similar meta-analyses show. Thus, the trade off of this technique would have been to increase the likelihood for a Type II error (falsely accepting the null hypothesis). Hence, no p value correction was done. Instead, the findings of this meta-analyses should be partly regarded as rather exploratory results (especially concerning the covariation between the variables). They need to be confirmed in future analyses. Some hypotheses, though, were explicated and theoretically based expectations described (see section 1.6).

[32] Post-hoc analyses revealed substantial and almost high effect sizes for child and parent trainings in developing countries (d = 0.74 and 0.76, respectively).

[33] About 3 500 statistical tests were conducted: $p\ value = \frac{0.05}{3500}$.

4.5.7 P Value Versus Effect Size

Regarding the conducted moderator analyses, some descriptively strongly diverging effect sizes were observable but did not reach a substantial level (e.g., the association of the trainer with the developer versus the author). Those results were not further commented. In the context of p value relativisation (e.g., Cohen, 1990; Greenwald, Gonzalez, Harris, & Guthrie, 1996; Kline, 2004; Sullivan & Feinn, 2012), this strategy needs to be seen critical. However, regarding all results, effect sizes and confidence intervals were reported and can be further analysed, also by other researchers.

4.6 Final Conclusion

The findings of this comprehensive and worldwide meta-analysis of high quality research evidence on psychosocial and educational preventive interventions for children and adolescents from low-SES supports the compelling evidence on the topic and adds surprising new conclusions. An unbiased and robust overall 15% improvement of the intervention compared to the control group could be found. Child and parent trainings can equally be applied to enhance children's educational development. Combined child-parent trainings can be regarded as promising approach but needs to be further investigated. Teacher trainings, on the other hand, showed no substantial influence on the children's development.

Effects on children's educational development seem to be relatively stable over time. Especially when implementing child trainings, the effects remain into adulthood. In contrast to that, parent training effectiveness seems to dissolve already 12 months after the programme ended. Thus, for achieving long-term improvements regarding the educational development, children should be approached directly.

Several factors could be found which influence the impact of child and parent trainings. When implementing child trainings, structuredness seems to be essential for actually achieving improvements regarding children's school development, whereas supervisions and observations seem to hinder the training efficacy more than they do good. In order to achieve higher improvements of children's basic as well as school development, implementation problems should be prevented and taken seriously when conducting child trainings. For improving basic and school develop-

ment through training the parents, the application of a manual is essential. School development can be further enhanced by training the executor. Basic development improvements are especially high if the parent training is conducted in a socially deprived area.

Despite the positive findings, if intending to close the developmental gap of low-SES children and adolescents, the present evidence is rather disillusioning. Hence, psychosocial and educational preventive interventions reduce the developmental gap associated with poverty but they are surely not able to close it and to equalise developmental chances for all children and adolescents regardless of their socio-economic status.

References

Abdi, H. (2007). The Bonferonni and Šidák corrections for multiple comparisons. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics.* Thousand Oaks, CA: SAGE Publications, Inc. Retrieved from http://www.utdallas.edu/\textasciitildeherve/Abdi-Bonferroni2007-pretty.pdf

Aber, J. L., Bennett, N. G., Conley, D. C., & Li, J. (1997). The effects of poverty on child health and development. *Annual Review of Public Health*, *18*(1), 463–483. doi: 10.1146/annurev.publhealth.18.1.463

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, *100*(2), 235–251. doi: 10.1037/0022-0663.100.2.235

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (1993). First-grade classroom behavior: Its short- and long-term consequences for school performance. *Child Development*, *64*(3), 801–804. doi: 10.2307/1131219

Alexander, K. L., Entwisle, D. R., & Thompson, M. S. (1987). School performance, status relations, and the structure of sentiment: Bringing the teacher back in. *American Sociological Review*, *52*(5), 665–682. doi: 10.2307/2095602

Antikainen, A., & Luukkainen, A. (2008). *Twenty-five years of educational reform initiatives in Finland*. Unpublished manuscript. Retrieved from http://wanda.uef.fi/\textasciitildeanti/publ/uudet/twenty\textunderscorefive\textunderscoreyears.pdf

Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, *54*(1), 517–545. doi: 10.1146/annurev.psych.54.111301.145442

Astuto, J., & Allen, L. (2009). *Home visitation and young children: An approach worth investing in? Social policy report. Volume xxiii, number iv*. Retrieved from ERIC database (ED509749).

Atkinson, A. B., & Marlier, E. (2010). *Analysing and measuring social inclusion in a global context*. New York, NY: United Nations. Retrieved from http://www.un.org/esa/socdev/publications/measuring-social-inclusion.pdf

Atkinson, T., Cantillon, B., Marlier, E., & Nolan, B. (2002). *Social indicators: The EU and social inclusion*. Oxford, England and New York, NY: Oxford University Press.

Autorengruppe Bildungsberichterstattung. (2010). *Bildung in Deutschland 2010: Ein indikatorengestützter Bericht mit einer Analyse zu Perspektiven des Bildungswesens im demografischen Wandel*. Bielefeld, Germany: W. Bertelsmann Verlag. Retrieved from http://www.bildungsbericht.de/daten2010/bb\textunderscore2010.pdf

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Hove, England and New York, NY: Psychology Press.

Badian, N. A. (1995). Predicting reading ability over the long term: The changing roles of letter naming, phonological awareness and orthographic processing. *Annals of Dyslexia*, *45*(1), 79–96. doi: 10.1007/BF02648213

Bakermans-Kranenburg, M. J., van IJzendoorn, Marinus H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin*, *129*(2), 195–215. doi: 10.1037/0033-2909.129.2.195

Barnett, W. S. (1992). Benefits of compensatory preschool education. *Journal of Human Resources*, *27*(2), 279–312. doi: 10.2307/145736

Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications*. Boulder, CO and Tempe, AZ: Education and the Public Interest Center & Education. Retrieved from http://nieer.org/resources/research/PreschoolLastingEffects.pdf

Baytop, C. M. (2006). Evaluating the effectiveness of programs to improve educational attainment of unwed african american teen mothers: A meta analysis. *The Journal of Negro Education*, *75*(3), 458–477.

Beelmann, A. (2006). Wirksamkeit von Präventionsmaßnahmen bei Kindern und Jugendlichen. *Zeitschrift für Klinische Psychologie und Psychotherapie*, *35*(2), 151–162. doi: 10.1026/1616-3443.35.2.151

Beelmann, A., & Lipsey, M. W. (in press). Meta-analysis of effect estimates from multiple studies. In M. W. Lipsey & D. S. Cordray (Eds.), *Field experimentation: Methods for evaluating what works, for whom, under what circumstances, how, and why.* Thousand Oaks: SAGE Publications.

Beelmann, A., Pfost, M., & Schmitt, C. (2014). Prävention und gesundheitsförderung bei kindern und jugendlichen. *Zeitschrift für Gesundheitspsychologie*, *22*(1), 1–14. doi: 10.1026/0943-8149/a000104

Beelmann, A., & Raabe, T. (2007). *Dissoziales Verhalten von Kindern und Jugendlichen: Erscheinungsformen, Entwicklung, Prävention und Intervention*. Göttingen, Germany: Hogrefe.

Beelmann, A., & Schmitt, C. (2012). Einflussfaktoren auf die Effektivität. In M. Fingerle & M. Grumm (Eds.), *Prävention von Verhaltensauffälligkeiten bei Kindern und Jugendlichen* (pp. 120–141). München, Germany and Basel, Switzerland: Ernst Reinhardt Verlag.

Blau, P. M., Duncan, O. D., & Tyree, A. (1978). *The American occupational structure*. New York, NY: Free Press.

Blishen, B. R. (1958). The construction and use of an occupational class scale. *The Canadian Journal of Economics and Political Science*, *24*(4), 519–531. doi: 10.2307/139088

Bolger, K. E., Patterson, C. J., Thompson, W. W., & Kupersmidt, J. B. (1995). Psychosocial adjustment among children experiencing persistent and intermittent family economic hardship. *Child Development*, *66*(4), 1107–1129. doi: 10.2307/1131802

Bollen, K. A., Glanville, J. L., & Stecklov, G. (2001). Socioeconomic status and class in studies of fertility and health in developing countries. *Annual Review of Sociology*, *27*(1), 153–185. doi: 10.1146/annurev.soc.27.1.153

Bollnow, O. F. (1963). Erziehung und Leben. *Erziehung und Leben IDE (Institute of Democratic*, *28*, 67–81. Retrieved from http://www.otto-friedrich-bollnow.de/doc/ErziehungundLeben4.pdf

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, England: John Wiley & Sons.

Bornstein, M. H., & Bradley, R. H. (2012). *Socioeconomic status, parenting, and child development*. New York, NY: Routledge.

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7th ed.). Berlin and Heidelberg, Germany: Springer-Verlag. Retrieved from http://link.springer.com/content/pdf/10.1007\%2F978-3-642-12770-0.pdf

Bos, W., Tarelli, I., Bremerich-Vos, A., & Schwippert, K. (2012). *Iglu 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster, Germany: Waxmann.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399. doi: 10.1146/annurev.psych.53.100901.135233

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the united states part i: Variations by age, ethnicity, and poverty status. *Child Development*, *72*(6), 1844–1867. doi: 10.1111/1467-8624.t01-1-00382

Bradshaw, J., & Finch, N. (2003). Overlaps in dimensions of poverty. *Journal of Social Policy*, *32*(4), 513–525. doi: 10.1017/S004727940300713X

Brandtstädter, J. (1982). Methodologische Grundfragen psychologischer Prävention. In J. Brandstädter & A. von Eye (Eds.), *Psychologische Prävention* (pp. 37–79). Bern: Huber.

Breitsameter, F., Hillenbrand, M., & Burgwitz, A. (2009). *Technik im Überblick: Die bedeutendsten Daten, Fakten, Ereignisse und Personen*. München, Germany:

Compact-Verlag.

Brockhaus. (2001). *Der Brockhaus Psychologie: Fühlen, Denken und Verhalten verstehen*. Mannheim and Leipzig, Germany: F.A. Brockhaus GmbH.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1989). Ecological systems theory. In R. Vasta (Ed.), *Annals of child development: Six theories of child development* (Vol. 6, pp. 187–249). Greenwich, CT: JAI Press.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children*, *7*(2), 55–71. doi: 10.2307/1602387

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology*, *99*(2), 353–395. doi: 10.1086/230268

Brooks-Gunn, J., & Markman, L. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *The Future of Children*, *15*(1), 139–168. doi: 10.1353/foc.2005.0001

Bruce, M. L. (1991). Poverty and psychiatric status: Longitudinal evidence from the New Haven epidemiologic catchment area study. *Archives of General Psychiatry*, *48*(5), 470–474. doi: 10.1001/archpsyc.1991.01810290082015

Burg, F. (2011). Zur Berechnung von Kaufkraftparitäten. *Statistisches Bundesamt, Wirtschaft und Statistik*, 793–801. Retrieved from https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Preise/Kaufkraftparitaeten\textunderscore82011.pdf?\textunderscore\textunderscoreblob=publicationFile

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*,

25(2), 140–165. doi: 10.1016/j.ecresq.2009.11.001

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *The Teachers College Record*, *112*(3), 579–620.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Comp. Retrieved from http://moodle.technion.ac.il/pluginfile.php/219643/mod\textunderscoreresource/content/0/Campbell\textunderscoreand\textunderscoreStanley\textunderscore1963.pdf

Campbell, S., & Stauffenberg, C. v. (2008). Child characteristics and family processes that predict behavioral readiness for school. In A. Booth & A. C. Crouter (Eds.), *Disparities in school readiness: How families contribute to transitions in school* (pp. 225–258). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Caplan, G. (1964). *Principles of preventive psychiatry*. New York, NY: Basic Books. Retrieved from https://books.google.de/books?id=qnIgAAAAMAAJ

Carrasco, M., Volkmar, F. R., & Bloch, M. H. (2012). Pharmacologic treatment of repetitive behaviors in autism spectrum disorders: Evidence of publication bias. *Pediatrics*, *129*(5), e1301-10. doi: 10.1542/peds.2011-3285

Cauthen, K., & Fass, S. (2007). *Measuring income and poverty in the United States.* Retrieved from http://www.nccp.org/publications/pdf/text\textunderscore707.pdf

Chambers, B., Cheung, A., Slavin, R. E., Smith, D., & Laurenzano, M. (2010). *Effective early childhood education programs: A systematic review*. Retrieved from ERIC database (ED527643).

Chao, R. K., & Willms, J. D. (2002). The effects of parenting practices on children's outcomes. In J. D. Willms (Ed.), *Vulnerable children: Findings from Canada's National Longitudinal Survey of Children and Youth* (pp. 149–165). Edmonton, Canada:

University of Alberta Press.

Chase-Lansdale, P. L., & Gordon, R. A. (1996). Economic hardship and the development of five–and six–year–olds: Neighborhood and regional perspectives. *Child Development*, *67*(6), 3338–3367. doi: 10.2307/1131782

Citro, C. F., & Michael, R. T. (1995). *Measuring poverty: A new approach*. Washington, DC: National Academy Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37–46. doi: 10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*(12), 1304–1312.

Coie, J. D., Watt, N. F., West, S. G., Hawkins, J. D., Asarnow, J. R., Markman, H. J., ... Long, B. (1993). The science of prevention. A conceptual framework and some directions for a national research program. *American Psychologist*, *48*, 1013–1022.

Coleman, H. A. (1940). The relationship of socio-economic status to the performance of junior high school students. *The Journal of Experimental Education*, *9*(1), 61–63.

Conger, R. D., Conger, K. J., & Elder, G. H. (1997). Family economic hardship and adolescent adjustment: Mediating and moderating processes. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 288–310). New York, NY: Russell Sage.

Cook, T. D., & Campbell, D. T. (Eds.). (1979a). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin Company.

Cook, T. D., & Campbell, D. T. (1979b). Validity. In T. D. Cook & D. T. Campbell (Eds.), *Quasi-experimentation* (pp. 37–94). Boston, MA: Houghton Mifflin Company.

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research*

synthesis and meta-analysis (2. ed ed.). New York, NY: Russell Sage Foundation.

Council of the European Union. (1975). Council decision of 22 july 1975 concerning a programme of pilot schemes and studies to combat poverty (75/458/eec). *Official Journal of the European Communities*, *L199*, 34–35. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31975D0458\&qid=1444305894880\&from=EN

Council of the European Union. (1985). Council decision of 19 december 1984 on specific community action to combat poverty (85/8/eec). *Official Journal of the European Communities*, *L2*, 24–25. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31985D0008\&from=EN

Crouse, G., Hauan, S., & Waters Rogers, A. (2008). *Indicators of welfare dependence: Annual report to congress*. Washington, DC: United States Department of Health and Human Services.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, *40*(2), 211–223. doi: 10.1017/S0033291709006114

Dahl, G., & Lochner, L. (2005). *The impact of family income on child achievement*. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w11279

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. doi: 10.1016/S0272-7358(97)00043-3

Darrow, C. L. (2009). *Language and literacy effects of curriculum interventions for preschools serving economically disadvantaged children: A meta analysis.* Retrieved from ERIC database (ED514949).

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, *19*(2), 294–304. doi: 10.1037/0893-3200

.19.2.294

Dawson, D. A. (1991). Family structure and children's health: United States, 1988 (PHS 91-1506). *Vital and Health Statistics. Series 10, Data from the National Health Survey*(178), 1–47.

Derzon, J. H., Sale, E., Springer, J. F., & Brounstein, P. (2005). Estimating intervention effectiveness: synthetic projection of field evaluation results. *The Journal of Primary Prevention*, *26*(4), 321–343. doi: 10.1007/s10935-005-5391-5

DESTATIS. (2015). *Explanatory notes: At-risk-of-poverty rate.* Wiesbaden, Germany. Retrieved from https://www.destatis.de/EN/Publications/STATmagazin/IncomeConsumptionLivingConditions/2012\textunderscore12/ExplanatoryNotes.html

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *The Journal of the American Medical Association*, *263*(10), 1385–1389. doi: 10.1001/jama.1990.03440100097014

Dickersin, K., Chan, S., Chalmersx, T. C., Sacks, H. S., & Smith, H. (1987). Publication bias and clinical trials. *Controlled Clinical Trials*, *8*(4), 343–353. doi: 10.1016/0197-2456(87)90155-3

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development*, *65*(2), 649–665. doi: 10.2307/1131407

Dornbusch, S. M., Ritter, P. L., & Steinberg, L. (1991). Community influences on the relation of family statuses to adolescent school performance: Differences between african americans and non-hispanic whites. *American journal of Education*, *99*(4), 543–567. doi: 10.1086/443997

Dorsch, F., Häcker, H. O., Stapf, K.-H., & Becker-Carus, C. (2009). *Dorsch psychologisches Wörterbuch* (15th ed.). Bern, CH: Huber. Retrieved from http://www.socialnet.de/rezensionen/isbn.php?isbn=978-3-456-84684-2

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American journal of community psychology*, *30*(2), 157–197.

Duncan, G., & Brooks-Gunn, J. (2000). Family poverty, welfare reform, and child development. *Child Development*, *71*(1), 188–196. doi: 10.1111/1467-8624.00133

Duncan, G., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early-childhood development. *Child Development*, *65*(2), 296–318. doi: 10.2307/1131385

Duncan, G., Yeung, W. J., Brooks-Gunn, J., & Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review*, *63*(3), 406–423. doi: 10.2307/2657556

Duncan, O. (1961). A socioeconomic index for all occupations. In A. J. Reiss (Ed.), *Occupations and social status* (pp. 109–138). New York, NY: Free Press.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology*, *41*(3-4), 327–350. doi: 10.1007/s10464-008-9165-0

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. doi: 10.1080/01621459.2000.10473905

Eamon, M. K. (2002). Effects of poverty on mathematics and reading achievement of young adolescents. *The Journal of Early Adolescence*, *22*(1), 49–74. doi: 10.1177/0272431602022001003

Early Childhood Learning and Knowledge Center. (2016). *Head Start program facts fiscal year 2015.* Retrieved from https://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/docs/head-start-fact-sheet-fy-2015.pdf

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis

detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. doi: 10.1136/bmj.315.7109.629

Elvik, R. (1998). Evaluating the statistical conclusion validity of weighted mean results in meta-analysis by analysing funnel graph diagrams. *Accident Analysis & Prevention*, *30*(2), 255–266. doi: 10.1016/S0001-4575(97)00076-6

Encyclopædia Britannica Online. (2015). *Poverty*. Chicago (World Headquarters), IL: Encyclopædia Britannica, Inc. Retrieved from http://www.britannica.com/topic/poverty

Erikson, E. H. (1963/1995). *Childhood and society* (Rev. ed.). London, England: Vintage.

European Commission. (2008). *Child poverty and well-being in the EU: Current status and way forward*. Luxembourg, Luxembourg: European Commission, Directorate-General for Employment, Social Affairs and Equal Opportunities, Unit E.2. Retrieved from http://ec.europa.eu/social/BlobServlet?docId=2049\&langId=en

European Commission. (2015). *Education and training monitor 2015 Finland*. Luxembourg, Luxembourg: Publications Office of the European Union. Retrieved from http://ec.europa.eu/education/tools/docs/2015/monitor2015-finland\textunderscoreen.pdf

Eurostat. (2016a). *At-risk-of poverty rate by detailed age group*. Luxembourg, Luxembourg: European Commission. Retrieved from Eurostat website (tessi120): http://ec.europa.eu/eurostat/data/database.

Eurostat. (2016b). *Income, saving and net lending/borrowing - current prices*. Luxembourg, Luxembourg: European Commission. Retrieved from Eurostat website (nama_inc_c): http://ec.europa.eu/eurostat/data/database.

Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist*, *59*(2), 77–92. doi: 10.1037/0003-066X.59.2.77

Farrington, D. P. (2003). Methodological quality standards for evaluation research. *The ANNALS of the American Academy of Political and Social Science*, *587*(1), 49–68. doi: 10.1177/0002716202250789

Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: A meta-analysis. *Australian and New Zealand Journal of Criminology*, *36*(2), 127–151. doi: 10.1375/acri.36.2.127

Federman, M., Garner, T. I., Short, K., Cutter, W. B., Kiely, J., Levine, D., ... McMillen, M. (1996). What does it mean to be poor in America? *Monthly Labor Review*, *119*(3), 3–17.

Felner, R. D., Brand, S., DuBois, D. L., Adan, A. M., Mulhall, P. F., & Evans, E. G. (1995). Socioeconomic disadvantage, proximal environmental experiences, and socioemotional and academic adjustment in early adolescence: Investigation of a mediated effects model. *Child Development*, *66*(3), 774–792. doi: 10.2307/1131950

FileMaker Inc. (2008). *FileMaker Pro (Version 10.0v1)*. Santa Clara, CA: FileMaker, Inc.

Fischer, W. (1982). *Armut in der Geschichte: Erscheinungsformen und Lösungsversuche der 'Sozialen Frage' in Europa seit dem Mittelalter*. Göttingen, Germany: Vandenhoeck & Ruprecht. Retrieved from http://digi20.digitale-sammlungen.de/de/fs1/object/display/bsb00047871\textunderscore00001.html

Galbraith, J. K. (1998). *The affluent society* (40th anniversary ed.). Boston, MA: Houghton Mifflin.

Garbarino, J. (1999). The effects of community violence on children. In L. Balter (Ed.), *Child psychology: A handbook of contemporary issues* (pp. 412–425). New York, NY: Psychology Press.

Garrett, P., Ng'andu, N., & Ferron, J. (1994). Poverty experiences of young children and the quality of their home environments. *Child Development*, *65*(2), 331–345. doi: 10.2307/1131387

Gilbert, D. L. (2011). *The American class structure in an age of growing inequality* (8th ed.). Los Angeles, CA: Pine Forge Press.

Gilbody, S. M., Song, F., Eastwood, A. J., & Sutton, A. (2000). The causes, consequences and detection of publication bias in psychiatry. *Acta Psychiatrica Scandinavica*, *102*(4), 241–249. doi: 10.1034/j.1600-0447.2000.102004241.x

Goodman, E. (1999). The role of socioeconomic status gradients in explaining differences in us adolescents' health. *American Journal of Public Health*, *89*(10), 1522–1528. doi: 10.2105/AJPH.89.10.1522

Greenwald, A. G., Gonzalez, R., Harris, R. J., & Guthrie, D. (1996). Effect sizes and *p* values: What should be reported and what should be replicated? *Psychophysiology*, *33*(2), 175–183. doi: 10.1111/j.1469-8986.1996.tb02121.x

Griffin, J. A. (2010). Research on the implementation of preschool intervention programs: Learning by doing. *Early Childhood Research Quarterly*, *25*(3), 267–269. doi: 10.1016/j.ecresq.2010.03.004

Günter, C. (Ed.). (1995). *Fachlexikon ABC psychologie* (5th ed.). Frankfurt am Main and Berlin, Germany: Harri Deutsch, Springer.

Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography*, *37*(4), 431–447.

Hagenaars, A., & de Vos, K. (1988). The definition and measurement of poverty. *The Journal of Human Resources*, *23*(2), 211–221. doi: 10.2307/145776

Hagenaars, A. J., Zaidi, M. A., & de Vos, K. (1994). *Poverty statistics in the late 1980s: Research based on micro-date*. Luxembourg, Luxembourg: Office for Official Publications of the European Communities (EUR-OP).

Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., … Dahlberg, L. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior: A systematic review. *American Journal of Preventive Medicine*, *33*(2 Suppl.), 114–129. doi: 10.1016/j.amepre.2007.04.012

Hanson, T. L., McLanahan, S. S., & Thomson, E. (1997). Economic resources, parental practices and child well-being. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 190–238). New York, NY: Russell Sage.

Hauser, R. M., & Huang, M.-H. (1997). Verbal ability and socioeconomic success: A trend analysis. *Social Science Research*, *26*(3), 331–376. doi: 10.1006/ssre.1997.0604

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107–128. doi: 10.3102/10769986006002107

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Heinrichs, N., Saßmann, H., Hahlweg, K., & Perrez, M. (2002). Prävention kindlicher Verhaltensstörungen. *Psychologische Rundschau*, *53*(4), 170–183. doi: 10.1026//0033-3042.53.4.170

Higgins, J. P. T., & Green, S. (2009). *Cochrane handbook for systematic reviews of interviews (Version 5.1.0)*. Cochrane Collaboration. Retrieved from www.handbook.cochrane.org

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. doi: 10.1136/bmj.327.7414.557

Hollingshead, A. B. (1975). *Four factor index of social status.* Unpublished manuscript. New Haven, CT: Yale University. Retrieved from http://psy6023.alliant.wikispaces.net/file/view/hollingshead+ses.pdf

Hollingshead, A. B., & Redlich, F. C. (1958). *Social class and mental illness: A community study*. New York, NY: John Wiley.

Hook, D. (2009). Erikson's Stages of Psychosocial Development. In J. Watts, K. Cockcroft, & N. Duncan (Eds.), *Developmental psychology* (pp. 283–313). Cape Town,

South Africa: UCT Press.

IBM Corporation. (2011). *IBM SPSS Statistics for Windows (Version 20.0)*. Armonk, NY: IBM Corp.

Indian Ministry of Law and Justice. (2013). The National Food Security Act. *The Gazette of India*, *Act. No. 20*. Retrieved from http://indiacode.nic.in/acts-in-pdf/202013.pdf

Ingersoll, R. M. (2002). *Out-of-field teaching, educational inequality, and the organization of schools: An exploratory analysis : A research report*. Seattle, WA: Center for the Study of Teaching and Policy. Retrieved from https://depts.washington.edu/ctpmail/PDFs/OutOfField-RI-01-2002.pdf

JARS Group. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist*, *63*(9), 839–851. doi: 10.1037/0003-066X.63.9.839

Johnson, S. M., & Birkeland, S. E. (2003). Pursuing a sense of success: New teachers explain their career decisions. *American Educational Research Journal*, *40*(3), 581–617. doi: 10.3102/00028312040003581

Johnson, S. M., Kardos, S. M., Kauffman, D., Liu, E., & Donaldson, M. L. (2004). The support gap: New teachers' early experiences in high-income and low-income schools. *Education Policy Analysis Archives*, *12*(61), 1–25. doi: 10.14507/epaa.v12n61.2004

Jordan, N., Huttenlocher, J., & Levine, S. C. (1992). Differential calculation abilities in young children from middle-and low-income families. *Developmental Psychology*, *28*(4), 644–653. doi: 10.1037/0012-1649.28.4.644

Kahl, J. A., & Davis, J. A. (1955). A comparison of indexes of socio-economic status. *American Sociological Review*, *20*(3), 317–325.

Kahnt, H. (2003). *Die Geschichte der Deutschen Mark in Ost und West*. Regenstauf, Germany: Gietl Verlag.

Kan, K., & Tsai, W.-D. (2005). Parenting practices and children's education outcomes. *Economics of Education Review*, *24*(1), 29–43. doi: 10.1016/j.econedurev.2004.04.006

Kauffman, D., Moore Johnson S., Kardos, S., Liu, E., & Peske, H. (2002). 'Lost at Sea': New teachers' experiences with curriculum and assessment. *The Teachers College Record*, *104*(2), 273–300. Retrieved from http://s3.amazonaws.com/academia.edu.documents/39076757/5433fef90cf294006f73430e.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA\&Expires=1467736083\&Signature=PzALdzeLkT9EFrhbyNJNlQVqIRQ\%3D\&response-content-disposition=inline\%3B\%20filename\%3DLost\textunderscoreat\textunderscoreSea\textunderscoreNew\textunderscoreTeachers\textunderscoreExperiences\textunderscorewit.pdf

Kesselring, T. (1999). *Jean Piaget* (2nd ed.). München, Germany: Verlag C.H. Beck.

Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8: A meta-analysis of classroom and home interventions. *Review of Educational Research*, *83*(3), 386–431. doi: 10.3102/0034654313483906

Klebanov, P. K., Brooks-Gunn, J., McCarton, C., & McCormick, M. C. (1998). The contribution of neighborhood and family income to developmental test scores over the first three years of life. *Child Development*, *69*(5), 1420–1436. doi: 10.2307/1132275

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research* (1st ed.). Washington, DC: American Psychological Association.

Korenman, S., Miller, J. E., & Sjaastad, J. E. (1995). Long-term poverty and child development in the United States: Results from the NLSY. *Children and Youth Services Review*, *17*(1-2), 127–155. doi: 10.1016/0190-7409(95)00006-X

Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of*

Educational Psychology, *93*(3), 451–455. doi: 10.1037/0022-0663.93.3.451

Lampert, T., & Kroll, L. E. (2006). Messung des sozioökonomischen Status in sozialepidemiologischen Studien. In M. Richter & K. Hurrelmann (Eds.), *Gesundheitliche Ungleichheit* (pp. 297–319). Wiesbaden: VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-90357-6\textunderscore18

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62. doi: 10.3102/01623737024001037

LaTeX Project Team. (2001). *LaTeX 2$_\epsilon$*. LaTeX Project Public License. Retrieved from http://www.latex-project.org/

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*(2), 275–313. doi: 10.3102/00346543076002275

Lee, V. E., Brooks-Gunn, J., & Schnur, E. (1988). Does Head Start work? A 1-year follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Developmental Psychology*, *24*(2), 210–222. doi: 10.1037/0012-1649.24.2.210

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Washington, DC: Economic Policy Institute.

Lelkes, O., & Gasior, K. (2012). *Income poverty and social exclusion in the EU: Situation in 2008 and trends (based on EU-SILC 2005-2009): Policy brief*. Vienna, Austria. Retrieved from http://www.euro.centre.org/data/1327061559\textunderscore78123.pdf

Lerner, R. M., & Castellino, D. R. (2002). Contemporary developmental theory and adolescence: Developmental systems and applied developmental science. *Journal of Adolescent Health*, *31*(6), 122–135. doi: 10.1016/S1054-139X(02)00495-0

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*, *126*(2), 309–337. doi: 10.1037//0033-2909.126.2.309

Lie, S., Linnakylä, P., & Roe, A. (2003). *Northern lights on PISA: Onity and diversity in the nordic countries in PISA 2000*. University of Oslo: Department of Teacher Education and School Development.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Vol. 49). Thousand Oaks, CA: SAGE Publications.

Lord, H., & Mahoney, J. L. (2007). Neighborhood crime and self-care: Risks for aggression and lower academic performance. *Developmental Psychology*, *43*(6), 1321–1333. doi: 10.1037/0012-1649.43.6.1321

Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *The ANNALS of the American Academy of Political and Social Science*, *587*(1), 84–109. doi: 10.1177/0002716202250793

Lugaila, T. A. (2003). *A child's day: 2000 (selected indicators of child well-being)* (Vol. 89). Washington, DC: US Census Bureau. Retrieved from http://www.census.gov/prod/2003pubs/p70-89.pdf

Mabry, J. H. (1997). Review of hart and risley's meaningful differences in the everyday experience of young american children. *The Behavior Analyst*, *20*(1), 25–30.

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, *20*(4), 641–654. doi: 10.1002/sim.698

MacLeod, J., & Nelson, G. (2000). Programs for the promotion of family wellness and the prevention of child maltreatment: A meta-analytic review. *Child Abuse & Neglect*, *24*(9), 1127–1149. doi: 10.1016/S0145-2134(00)00178-2

Mahoney, J., & Eccles, J. (2008). Organized activity participation for children from

low- and middle-income families. In A. Booth & A. C. Crouter (Eds.), *Disparities in school readiness: How families contribute to transitions in school* (pp. 207–222). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Manning, M., Homel, R., & Smith, C. (2010). A meta-analysis of the effects of early developmental prevention programs in at-risk populations on non-health outcomes in adolescence. *Children and Youth Services Review*, *32*(4), 506–519. doi: 10.1016/j.childyouth.2009.11.003

Mayer, S. E. (2002). *The influence of parental income on children's outcomes*. Wellington, NZ: Knowledge Management Group, Ministry of Social Development.

McCormick, C., & Mason, J. M. (1984). *Intervention procedures for increasing preschool children's interest in and knowledge about reading (Technical report No. 312)*. Champaign, IL: University of Illinois at Urbana-Champaign. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/18043/ctrstreadtechrepv01984i00312\textunderscoreopt.pdf?sequence=1

McCoy, M. G., Frick, P. J., Loney, B. R., & Ellis, M. L. (1999). The potential mediating role of parenting practices in the development of conduct problems in a clinic-referred sample. *Journal of Child and Family Studies*, *8*(4), 477–494. doi: 10.1023/A:1021907905277

McGuire, C., & White, G. D. (1955). *The measurement of social status*. Research paper in Human Development No. 3 (rev.). Austin, TX: Deptartment of Educational Psychology, University of Texas.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, *53*(2), 185–204. doi: 10.1037/0003-066X.53.2.185

Mercy, J. A., & Steelman, L. C. (1982). Familial influence on the intellectual attainment of children. *American Sociological Review*, *47*(4), 532–542. doi: 10.2307/2095197

Microsoft. (2008). *Microsoft Office Excel 2007 (Version 12.0.6749.5000)*. Redmond, WA: Microsoft Corporation.

Miller, D. C., & Salkind, N. J. (2002). *Handbook of research design and social measurement* (6th ed.). Thousand Oaks, CA: SAGE Publications.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2009). Risk factors for learning-related behavior problems at 24 months of age: Population-based estimates. *Journal of Abnormal Child Psychology*, *37*(3), 401–413. doi: 10.1007/s10802-008-9279-8

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, *11*(2), 364–386. doi: 10.1177/1094428106291059

Mortensen, J. A., & Mastergeorge, A. M. (2014). A meta-analytic review of relationship-based interventions for low-income families with infants and toddlers: Facilitating supportive parent-child interactions. *Infant Mental Health Journal*, *35*(4), 336–353. doi: 10.1002/imhj.21451

Mpofu, E., & van de Vijver, F. J. (2000). Taxonomic structure in early to middle childhood: A longitudinal study with Zimbabwean schoolchildren. *International Journal of Behavioral Development*, *24*(2), 204–212. doi: 10.1080/016502500383331

Murtin, F., & d'Ercole, M. M. (2015). Household wealth inequality across OECD countries: New OECD evidence. *OECD Statistics Brief*, *21*, 1–8. Retrieved from http://www.oecd.org/std/household-wealth-inequality-across-OECD-countries-OECDSB21.pdf

Nam, C. B., & Boyd, M. (2004). Occupational status in 2000; over a century of census-based measurement. *Population Research and Policy Review*, *23*(4), 327–358. doi: 10.1023/B:POPU.0000040045.51228.34

Nam, C. B., & Powers, M. G. (1983). *The socioeconomic approach to status measurement: With a guide to occupational and socioeconomic status scores*. Houston, TX: Cap and Gown Press.

National Science Foundation. (2003). *Women, minorities, and persons with disabilities in science and engineering: 2002*. Retrieved from NSF website (NSF 03-312):

http://www.nsf.gov/statistics/nsf03312/pdf/women02.pdf.

Nelson, G., Laurendeau, M.-C., Chamberland, C., & Peirson, L. (2001). A review and analysis of programs to promote family wellness and prevent the maltreatment of pre-school and elementary school-aged children. In I. Prilleltensky, G. Nelson, & L. Peirson (Eds.), *Promoting family wellness and preventing child maltreatment: Fundamantals for thinking and action* (pp. 220–272). Toronto, Canada: University of Toronto Press.

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment*, *6*(31), No pagination specified. doi: 10.1037/1522-3736.6.1.631a

Nolan, B., & Whelan, C. T. (1996). *Resources, deprivation, and poverty*. Oxford, England and New York, NY: Clarendon Press and Oxford University Press.

OECD. (2012). *OECD education statistics*. Paris, France: OECD Publishing. doi: 10.1787/edu-db-data-en

OECD. (2013). *Education policy outlook Finland*. Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/edu/EDUCATION\%20POLICY\%20OUTLOOK\%20FINLAND\textunderscoreEN.pdf

OECD. (2013). *What are equivalence scales?* Paris, France: OECD Project on Income Distribution and Poverty. Retrieved from http://www.oecd.org/eco/growth/OECD-Note-EquivalenceScales.pdf

OECD. (2014). *Education at a glance 2014. OECD indicators*. Paris, France: OECD Publishing. doi: 10.1787/eag-2014-en

OECD. (2015). *Education at a glance 2015*. Paris, France: OECD Publishing. doi: 10.1787/eag-2015-en

OECD. (2016a). *Gross domestic product (GDP) (indicator)*. doi: 10.1787/dc2f7aec-en

OECD. (2016b). *Poverty rate (indicator)*. doi: 10.1787/0fe1315d-en

Ortega, S. T., & Corzine, J. (1990). Socioeconomic status and mental disorders. In J. R. Greenley (Ed.), *Research in community and mental health: A research annual: Mental disorder in social context* (pp. 149–182). Greenwich, CT: JAI Press.

Parcel, T. L., & Menaghan, E. G. (1990). Maternal working conditions and children's verbal facility: Studying the intergenerational transmission of inequality from mothers to young children. *Social Psychology Quarterly*, *53*(2), 132–147. doi: 10.2307/2786675

Perrez, M. (1994). Optimierung und Prävention im erzieherischen Bereich. In K. A. Schneewind (Ed.), *Enzyklopädie der Psychologie. Psychologie der Erziehung und Sozialisation. Pädagogische Psychologie* (Vol. 1, pp. 585–617). Göttingen: Hogrefe.

Perrez, M., & Baumann, U. (2011). Systematik der klinisch-psychologischen Intervention. In U. Baumann & M. Perrez (Eds.), *Lehrbuch Klinische Psychologie - Psychotherapie* (pp. 341–351). Bern, Switzerland: Huber.

Perry, B. (2002). The mismatch between income measures and direct outcome measures of poverty. *Social Policy Journal of New Zealand*, *19*, 101–127. Retrieved from https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/journals-and-magazines/social-policy-journal/spj19/19-pages-101-127.pdf

Phillips, D. A., Voran, M., Kisker, E., Howes, C., & Whitebook, M. (1994). Child care for children in poverty: Opportunity or inequity? *Child Development*, *65*(2), 472–492. doi: 10.2307/1131397

Pigott, T. D. (2009). Methods for handling missing data in research synthesis. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 399–415). New York, NY: Russell Sage Foundation.

Pilon, M. (2009). *The buying power of a dollar, on a downswing.* Retrieved from http://blogs.wsj.com/wallet/2009/01/28/the-buying-power-of-a-dollar-on-a-downswing/?mod=googlenews\textunderscorewsj

Pollert, A., Kirchner, B., & Polzin, J. M. (2013). *Duden Wirtschaft von A bis Z: Grundlagenwissen für Schule und Studium, Beruf und Alltag* (5th ed.). Berlin, Germany: Bibliographisches Institut GmbH. Retrieved from http://www.bpb.de/nachschlagen/lexika/lexikon-der-wirtschaft/19157/einkommen

Prenzel, M., Sälzer, C., Klieme, E., & Köller, O. (2013). *PISA 2012: Fortschritte und Herausforderungen in Deutschland*. Münster, Germany: Waxmann.

Puma, M., Bell, S., Cook, R., Heid, C., & Lopez, M. (2005). *Head Start impact study. First year findings*. Administration for Children & Families. Retrieved from ERIC database (ED543015).

Randolph, J. J., & Edmondson, R. S. (2005). Using the binomial effect size display (BESD) to present the magnitude of effect sizes to the evaluation audience. *Practical Assessment Research & Evaluation*, *10*(14), 1–7. Retrieved from http://pareonline.net/getvn.asp?v=10\&n=14

Ravallion, M., Chen, S., & Sangraula, P. (2009). Dollar a day revisited. *The World Bank Economic Review*, *23*(2), 163–184. doi: 10.1093/wber/lhp007

Rigby, S. H. (1987). *Marxism and history: A critical introduction*. Manchester, England: Manchester University Press.

Rist, R. C. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review*, *40*(3), 411–451.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. doi: 10.1037/0033-2909.86.3.638

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*(2), 166–169. doi: 10.1037/0022-0663.74.2.166

Rothstein, H., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, England and Hoboken, NJ: Wiley.

Sabates, R. (2008). *The impact of lifelong learning on poverty reduction: IFLL public value paper 1*. Leicester, England: National Institute of Adult and Continuing Education (NIACE) publications. Retrieved from http://www.niace.org.uk/lifelonglearninginquiry/docs/Public-value-paper-1.pdf

Sampson, R. J., & Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *American Journal of Sociology*, *94*(4), 774–802. doi: 10.1086/229068

Saxe, G. B., Guberman, S. R., & Gearhart, M. (1987). Social processes in early number development. *Monographs of the Society for Research in Child Development*, *52*(2), 1–162. doi: 10.2307/1166071

Scottish Poverty Information Unit. (2011). *Definitions of poverty*. Retrieved from http://www.bbc.co.uk/scotland/education/int/ms/health/wealth/def\textunderscoreof\textunderscorepoverty/definitions.shtml

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shields, M. K., & Behrman, R. E. (2000). Children and computer technology: Analysis and recommendations. *The Future of Children*, *10*(2), 4–30. doi: 10.2307/1602687

Simpson, K. M., Porter, K., McConnell, E. S., Colón-Emeric, C., Daily, K. A., Stalzer, A., & Anderson, R. A. (2013). Tool for evaluating research implementation challenges: A sense-making protocol for addressing implementation challenges in complex research settings. *Implementation Science*, *8*(2), 1–12. doi: 10.1186/1748-5908-8-2

Slavin, R. E., Lake, C., Cheung, A., & Davis, S. (2009). *Beyond the basics: Effective reading programs for the upper elementary grades*. Center for Data-Driven Refrom in Education. Retrieved from ERIC database (ED527574).

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*,

79(2), 839–911. doi: 10.3102/0034654308330968

Smith, J. D., Schneider, B. H., Smith, P. K., & Ananiadou, K. (2004). The effectiveness of whole-school antibullying programs: A synthesis of evaluation research. *School Psychology Review*, *33*(4), 547–560. doi: 10.1080/15564880701263155

Smith, J. R., Brooks-Gunn, J., & Klebanov, P. K. (1997). The consequences of living in poverty for young children's cognitive and verbal ability and early school achievement. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 132–189). New York, NY: Russell Sage.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance–or vice versa. *Journal of the American Statistical Association*, *54*(285), 30–34. doi: 10.2307/2282137

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development*, *57*(3), 646–659. doi: 10.2307/1130343

Stipek, D. J., & Ryan, R. H. (1997). Economically disadvantaged preschoolers: Ready to learn but further to go. *Developmental Psychology*, *33*(4), 711–723. doi: 10.1037/0012-1649.33.4.711

Sullivan, G. M., & Feinn, R. (2012). Using effect size - or why the p value is not enough. *Journal of Graduate Medical Education*, *4*(3), 279–282. doi: 10.4300/JGME-D-12-00156.1

Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? a meta-analytic review of home visiting programs for families with young children. *Child Development*, *75*(5), 1435–1456. doi: 10.1111/j.1467-8624.2004.00750.x

Swiss Academic Software. (2013). *Citavi (Version 4.4.0.28)*. Wädenswil, Switzerland: Swiss Academic Software GmbH.

The Canberra Group. (2001). *Final report and recommendations*. Ottawa, Canada: Can-

berra Group. Retrieved from http://www.unece.org/fileadmin/DAM/stats/groups/cgh/Canberra\%20Group\%20Handbook\%20(2001).pdf

The R Core Group. (2015). *R Wooden Christmas-Tree (Version 3.2.3)*. Vienna, Austria: The R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Thornton, A. (2000). Publication bias in meta-analysis its causes and consequences. *Journal of Clinical Epidemiology*, *53*(2), 207–216. doi: 10.1016/S0895-4356(99)00161-4

Tobler, N. S. (1986). Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results of program participants compared to a control or comparison group. *Journal of Drug Issues*, *16*(4), 537–567. doi: 10.1177/002204268601600405

Townsend, P. (1979). *Poverty in the united kingdom: A survey of household resources and standards of living*. London, England: Allen Lande and Penguin Books. Retrieved from http://poverty.ac.uk/sites/poverty/files/townsend-book-pdfs/PIUK/piuk-whole.pdf

UN General Assembly. (2015). *Transforming our world: The 2030 agenda for sustainable development (a/70/l.1)*. Retrieved from http://www.un.org/ga/search/view\textunderscoredoc.asp?symbol=A/RES/70/1\&Lang=E

UNICEF. (2000). *A league table of child poverty in rich nations* (Vol. 1). Florence, Italy: UNICEF Innocenti Research Centre.

United Nations. (1990). *Convention on the rights of the child. Article 27(1)*. Geneva, Switzerland: United Nations Office of the High Commissioner for Human Rights. Retrieved from http://www.ohchr.org/EN/ProfessionalInterest/Pages/CRC.aspx

United Nations. (1995). *Report of the world summit for social development (A/CONF.166/9, Chapter II, para. 19)*. Copenhagen, Denmark: United Nations World Summit for Social Development. Retrieved from http://www.un.org/

documents/ga/conf166/aconf166-9.htm

United Nations. (1998). *Integrated and coordinated implementation of and follow-up to major United Nations conferences and summits (E/1998/73)*. New York, NY: United Nations Economic and Social Council. Retrieved from https://www.unsceb.org/CEBPublicFiles/press/9818151e.pdf

United Nations. (2001). *Substantive issues arising in the implementation of the international covenant on economic, social and cultural rights: Poverty and the international covenant on economic, social and cultural rights (E/C.12/2001/10, p. 2–3)*. Geneva, Switzerland: United Nations Committee on Economic, Social and Cultural Rights. Retrieved from http://www2.ohchr.org/english/bodies/cescr/docs/statements/E.C.12.2001.10Poverty-2001.pdf

United Nations. (2006). *The millennium development goals report*. New York, NY: United Nations Department of Economic and Social Affairs. Retrieved from http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2006/MDGReport2006.pdf

United Nations. (2014). *Human development report 2014: Sustaining human progress: Reducing vulnerability and building resilience*. New York, NY: United Nations Development Programme. Retrieved from http://hdr.undp.org/sites/default/files/hdr14-report-en-1.pdf

United States Congress. (1964). *Economic Opportunity Act of 1964 (public law 88-452)*. Retrieved from http://www.gpo.gov/fdsys/pkg/STATUTE-78/pdf/STATUTE-78-Pg508.pdf

Valsiner, J., & Connolly, K. J. (2003). *Handbook of developmental psychology*. London, England and Thousand Oaks, CA: SAGE Publications.

Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Journal of Psychology*, *215*(2), 104–121. doi: 10.1027/0044-3409.215.2.104

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package.

Journal of Statistical Software, *36*(3), 1–48. Retrieved from http://www.jstatsoft.org/v36/i03/

Vygotsky, L. S. (1978/1997). Interaction between learning and development. In M. Gauvain & M. Cole (Eds.), *Readings on the development of children* (pp. 79–91). New York, NY: W. H. Freeman and Company.

Warner, W. L., Meeken, M., & Eells, K. (1960). *Social class in America. A manual of procedures for measurent of social status*. New York, NY: Harper & Brothers.

West, J., Denton, K., & Germino-Hausken, E. (2000). *America's kindergartners: Findings from the early childhood longitudinal study, kindergarten class of 1998-99, fall 1998 (NCES 2000-070)*. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2000/2000070.pdf

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, *69*(3), 848–872. doi: 10.1111/j.1467-8624.1998.tb06247.x

Wilson, D. B. (2005). *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved from http://mason.gmu.edu/\textasciitildedwilsonb/ma.html

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *71*(1), 136–149. doi: 10.1037/0022-006X.71.1.136

Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public school pre-kindergarten programs. *Early Childhood Research Quarterly*, *23*(3), 314–329. doi: 10.1016/j.ecresq.2008.02.003

World Bank. (1990). *World development report: Poverty*. New York, NY: Oxford University Press. Retrieved from https://openknowledge.worldbank.org/bitstream/handle/10986/5973/WDR\%201990\%20-\%20English.pdf?sequence=5

World Health Organization. (2015). *World health statistics 2015*. Geneva, Switzerland: World Health Organization. Retrieved from http://gbv.eblib.com/patron/

FullRecord.aspx?p=2070501

Yeung, W. J., Linver, M. R., & Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development*, *73*(6), 1861–1879. doi: 10.1111/1467-8624.t01-1-00511

Yoshikawa, H., Aber, J. L., & Beardslee, W. R. (2012). The effects of poverty on the mental, emotional, and behavioral health of children and youth: Implications for prevention. *American Psychologist*, *67*(4), 272–284. doi: 10.1037/a0028015

Zigler, E., Taussig, C., & Black, K. (1992). Early childhood intervention: A promising preventative for juvenile delinquency. *American Psychologist*, *47*(8), 997–1006. doi: 10.1037/0003-066X.47.8.997

Zill, N., Resnick, G., Kim, K., O'Donnell, K., Sorongon, A., Ziv, Y., ... others (2006). *Head Start performance measures center family and child experiences survey (faces 2000): Technical report*. Washington, DC: US Department of Health and Human Services. Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/tech2k\textunderscorefinal2.pdf

Appendices

A Calculating the Low-Income Thresholds

In this appendix a detailed description will be given on how the low-income thresholds were calculated. As mentioned already in section 1.2.2, the aim was to obtain thresholds which are comparable across countries worldwide and over time since no such database exists. Before the calculations could begin, several assumption had to be made concerning the income database, the low-income-SES threshold, the anchor currency, and the time intervals

Setting the Assumptions

Income Database. Even if a definition of income would be set, such as "...all earnings in form of money or property which a person, a household, or a company obtains in a certain period of time"[34], there is still a range of possible options what kind of income to choose. For instance, it needs to be decided whether the transfer income (e.g., including unemployment or child benefits) is considered or whether the income is adjusted for inflation or taxes. Consequently, the type of income which organisations and institutions assess differs substantially. The Organisation for Economic Cooperation and Development (OECD), for instance, organises one of the most extensive worldwide databases for the *disposable income*, which is "...derived from total income [gross income] by deducting current transfers paid" (The Canberra Group, 2001, p. 22) and is reported per capita in US Dollars. Likewise, the Directorate-General of the European Commission for Statistical Information (Eurostat) collects the *real disposable income of households* broken down per capita in purchasing power standards (PPS). PPS is an artificial currency which takes the purchasing power of each member states of the European Union into account and by that, makes a comparison possible (Burg, 2011). Both databases were suited for serving as core database for the calculation of the low-income thresholds. However, the OECD-database dates back to the year 1960 whereas Eurostat only offers incomplete data starting in 1995. Moreover, the OECD-database contains income data on 39 states worldwide, compared to 32 countries Europe-wide in the Eurostat-database. Thus, the OECD-database was used as income database.

[34] Definition from (Pollert, Kirchner, & Polzin, 2013)

Low-Income-SES Threshold. In the United States the official low-income threshold has been developed 40 years ago. It was based on food consumption which, then, accounted for one-third of the income. Hereafter, *poor* is who earns less than the average cost for food multiplied by the factor three (Cauthen & Fass, 2007). This indicator has many drawbacks. It is rather an absolute measure of low-SES since the general living standard within the country was not taken into account. It is also based on outdated assumptions because the expenditures have changed. For instance, family expenditures, such as child care, have grown disproportionally to the income since the 1960s (Cauthen & Fass, 2007). Even if the index would have been adjusted for inflation, the low-income line applied by the United States should rather be interpreted as a measure of severe misery than of relative material deprivation. It was therefore not suitable as threshold for the meta-analysis at hand.

Since about 20 years the European Commission uses a relative poverty threshold which has become a conventional and accepted method over the years applied within and outside Europe (Bradshaw & Finch, 2003): 60% of the national median equivalised net income. The equivalised net income "...is a per capita [disposable] income per household member weighted in proportion to the member's needs" (DESTATIS, 2015, para. 6). The term *equvalised* will be explained in more detail at the end of this appendix (see paragraph ***Weighting the Thresholds***). The mentioned 60%-threshold is collected by Eurostat and can be interpreted as an indicator for (risk of) financial poverty and thereby fits the definition in section 1.2.1. With this measure the wealth of each country can be taken into account. Furthermore, it can be judged "...practicable and immediately applicable..." (T. Atkinson et al., 2002, p. 92) because it only requires income information of the population to calculate it. It is available in purchasing power standards allowing a comparison between countries inside Europe and throughout the world. Considering further the timeliness of the threshold due to its self-updating property (T. Atkinson et al., 2002), it was regarded as suitable for the purpose of this meta-analysis.

Anchor Currency. There exist more than 160 official currencies worldwide[35] which makes it hard to compare studies from countries which use different currencies. Ad-

[35] Retrieved from http://www.iso.org/iso/, ISO 4217:2008

ditionally, change of currencies within a country are not seldom as the German history shows: The currency changed five times in the 20th century from the *Mark* to the *Rentenmark* to the *Reichsmark* and the *Alliierte Militärmark*, to the *Deutsche Mark* or the *Mark* in the German Democratic Republic, respectively, and finally to the *Euro* (Kahnt, 2003). From an economic point of view (i.e., considering the time it would have taken to calculate low-income thresholds for individual currencies or even several currencies within a country), it was decided to use one anchor currency. The US Dollar was chosen as such for this meta-analysis since it is internationally accepted and widely used in the area of foreign currency dealings and thus already implemented as anchor currency.

Time Intervals. Currencies are subject to fluctuations which is attributable to changes in demand and supply, the interest policy of a country, inflation processes, or legislative regulations. The purchasing power of the Dollar has fallen since 1913 by 95% (Pilon, 2009). In contrast to that, the value of the Euro has increased by around 40% between its introduction in 2002 until the year 2008 compared to the US Dollar. Since the financial crises in 2008 the value fell by 30% until 2015[36]. Since fluctuations in the value of currencies influences the income, they need to be considered when calculating low-income thresholds. Estimating an annual threshold for every country available would be uneconomical. In order to consider the fluctuations, though, four time intervals were introduced and, depending on the data available, for each country up to four low-income thresholds were calculated.

In a nutshell, the following assumptions were made for the calculation of the low-income thresholds:

I. **Income Database**: Table "Disposable income per capita" provided by the OECD[37]

II. **Low-income-SES Threshold**: Table "At-risk-of-poverty threshold in PPS" pro-

[36] Retrieved from http://www.ecb.europa.eu/stats/exchange/eurofxref/html/eurofxref-graph-usd.en.html

[37] Retrieved from http://www.oecd.org → Statistics → Disposable Income → "Net national income per capita, US Dollar, current prices, current PPPs"; now available with minor changes under doi: 10.1787/af9be38a-en.

vided by Eurostat[38]

III. **Currency**: US Dollar

IV. **Time Intervals**: Four time intervals "until 1979", "from 1980 to 1989", "from 1990 to 1999", "from 2000 to 2011"

Calculating the Low-Income-SES Thresholds

In order to calculate the low-income thresholds, two steps were necessary: First, the limited "At-risk-of-poverty threshold" table (*threshold table*) was structurally adapted to the more extensive "Disposable income per capita" table (*income table*). Then the low-income estimators and thresholds were calculated. Subsequent to those paragraphs, alternative sources will be listed for the countries uncovered by the above described procedure. Finally, an equivalence scale will be introduced in order to take into account different household constellations in the primary studies.

Adaption of the Tables. In order to apply the low-income thresholds of the threshold table on the more extensive data of the income table, the threshold table had to be converted from Euro to US Dollar. This was done for all the 32 countries starting with the year 2002 (introduction of the Euro) until the year 2010. Then the countries which were represented in both tables (threshold and income table) were extracted, resulting in the following 24 countries: Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany (including the former GDR since 1991), Greece, Hungary, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. In a next step, the income table was reduced to those countries and the above time span from 2002 until 2010.

Calculating Estimators and Thresholds. If all necessary data was available for the respective years (2002-2010), the proportion of the threshold table to the income table

[38] Retrieved from http://epp.eurostat.ec.europa.eu → data → under "Tables by themes": open Population and social conditions → Income and living conditions → Income distribution and monetary poverty → Monetary poverty → At-risk-of-poverty threshold (source: SILC) http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&language=en&pcode=tessi014&plugin=1 TEST

was calculated for each of the above listed countries. This resulted in a new table (estimator table) which then contained between three and nine percentage values per country. For each country the average over the years was calculated (country-specific estimator, CSE). The CSE was then averaged over all countries leading to the transnational estimator (TNE). For calculating the low-income thresholds, the income table needed to be reduced to the four specified time intervals by averaging the incomes for each country within one interval across the included years: until 1979, from 1980 to 1989, from 1990 to 1999, and from 2000 to 2011. The two estimators CSE and TNE were applied on those averaged disposable incomes. The TNE was only used if not enough data was available to obtain the CSE. For all the countries listed under paragraph ***Adaptation of the Tables***, a CSE estimator could be calculated. For the other countries, the TNE was applied (i.e., Austria, Canada, Chile, India, Indonesia, Israel, Japan, Korea, Mexico, New Zealand, Russia, South Africa, and the United States of America). Some countries had data available in the threshold table but not in the income table, so none of the two estimators could be used. In order to still obtain a low-income threshold for them, the data of the threshold table from 2002 until 2010 was averaged for each of those countries and then transferred to US Dollars. This data then served as low-income thresholds. That applied to the following countries: Bulgaria, Croatia, Latvia, Lithuania, Malta, Romania, Turkey, and Cyprus. With these strategies a table was merged containing the thresholds of 45 countries, at best for all the four time intervals. This table can be found in appendix B.

Alternative Sources. For countries which were neither represented in the threshold table nor in the income table, an alternative approach was needed. It was presumed that the majority of the potential primary studies were conducted in countries listed in appendix B. Hence, no low-income thresholds were calculated for the remaining countries due to economic reasons. Instead, the following list of sources was collected so that the threshold for a particular country could be calculated if necessary. In those cases the estimator of choice was the TNE.

- **World Income Inequality Database** by the UNU−WIDER[39]
  (Available from: http://www.wider.unu.edu/research/Database/en_GB/wiid/)

[39] The United Nations University World Institute for Development Economics Research

- **The World Top Income Database** by the Paris School of Economics
  (Available from: http://g-mond.parisschoolofeconomics.eu/topincomes/)
- **Country and Lending Groups by Income** by the World Bank
  (Available from: http://data.worldbank.org/about/country-classifications/country-and-lending-groups#Low_income)

Weighting the Thresholds. Different household compositions need unequal household incomes for acquiring the same standard of living. Those differences are important to consider, since the primary studies will analyse diverse populations, such as including exclusively single mothers or married couples. So-called equivalence scales standardise the size (single mother with one child versus two parents with two children) and the structure (adult versus child) of the economic resources (here: household income). This has the advantage of taking into account heterogeneous needs (e.g., food for a baby versus for an adult) as well as savings due to a common household (e.g., common washing machine). A first scale has been introduced by the OECD in 1982 which was then modified by A. J. Hagenaars, Zaidi, and de Vos in 1994 and after that adopted by Eurostat in the late 1990s (OECD, 2013). This so-called "OECD-modified equivalence scale" was applied in for the meta-analysis at hand. It assigns the first adult person of a household a weight of 1.0. Each following adult is weighted with 0.5 and each child (under the age of 14) is assigned a weight of 0.3. Adding up those weights yields to a value which, multiplied with the low-income threshold of the respective country in a specific year, results in a household size adapted threshold. A practical example of the scale will be given in appendix B.

B Low-Income Thresholds per Capita and Year in US Dollar

The following table is the result of the calculations described in appendix A. It shows the low-income thresholds for 45 countries and four time intervals, together with the underlying percentage estimator (i.e., *Estimator*). For the country-specific estimator (CSE), the exact percentages which were applied to the income table are depicted. The transnational estimator (TNE which is the average of the CSEs) is 42% and is the same for each country which is represented in the income table but not in the threshold table. For those countries with available data in the threshold table but not in the income table, the threshold was taken directly from the former (transferred to US Dollars) and is therefore termed *direct* in the table.

Example for Applying the Equivalence Scale. As mentioned at the end of appendix A, an example will be described here for the application of the equivalence scale. Assuming that an international research team conducted a study in the Netherlands and published it in 1985. The study fits the inclusion criteria; The last question remaining is whether the sample is materially deprived. The authors reported that the majority of the analysed sample families is composed of the parents and two minor children with an average age of six years. Information on the income is also given: More than 70% of the sample has an average household income of 16 500Fl. (Dutch guilder). Transforming this amount would yield 8 500US$ (exchange rate in 1985: 1US$ = 1.95Fl.). The equivalence weight would be calculated as follows: 1.0 + 0.5 + 0.3 + 0.3 = 2.1 (first adult, second adult, first child, second child). Taking the low-income threshold per capita in the Netherlands in 1985 (4 817$) and multiplying it with the equivalence weight of 2.1, results in a low-income threshold of 10 116$ which is higher than 8 500$. Hence, the study would be included in the meta-analysis.

	Low-income thresholds in US $				
Country	Until 1979	1980-89	1990-99	2000-2011	Estimator (%)
Australia	2 330	4 830	7 445	11 973	TNE
Austria	2 559	5 788	9 388	14 112	48

Continued on the next page...

	Low-income thresholds in US $				
Country	Until 1979	1980-89	1990-99	2000-2011	Estimator (%)
Belgium	2 526	5 467	8 966	13 313	47
Bulgaria	n.a.	n.a.	n.a.	1 436	direct
Canada	2 489	5 404	7 992	12 395	TNE
Chile	n.a.	n.a.	3 090	4 158	TNE
Croatia	n.a.	n.a.	n.a.	3 655	direct
Cyprus	n.a.	n.a.	n.a.	12 330	direct
Czech Republic	n.a.	n.a.	2 679	4 200	25
Denmark	3 282	7 030	11 141	17 199	60
Estonia	n.a.	n.a.	1 383	2 859	21
Finland	2 190	5 205	7 664	13 674	50
France	2 408	5 237	8 348	12 504	47
Germany	2 438	5 419	8 791	12 994	47
Greece	1 599	3 147	4 649	7 453	35
Hungary	n.a.	n.a.	1 601	2 813	21
Iceland	3 969	9 732	15 063	20 137	76
India	n.a.	n.a.	n.a.	1 019	TNE
Indonesia	n.a.	n.a.	n.a.	1 278	TNE
Ireland	1 788	3 753	7 575	14 869	53
Israel	n.a.	n.a.	7 281	8 926	TNE
Italy	2 170	5 004	8 175	11 518	46
Japan	1 827	4 526	7 731	10 268	TNE
Korea	491	1 696	4 467	8 501	TNE
Latvia	n.a.	n.a.	n.a.	3 144	direct
Lithuania	n.a.	n.a.	n.a.	2 862	direct
Luxembourg	2 440	7 080	14 243	22 076	46
Malta	n.a.	n.a.	n.a.	7 861	direct
Mexico	1 043	2 098	2 973	4 770	TNE
Netherlands	2 287	4 817	8 027	13 650	43
New Zealand	2 066	4 022	5 746	8 758	TNE

Continued on the next page...

	Low-income thresholds in US $				
Country	Until 1979	1980-89	1990-99	2000-2011	Estimator (%)
Norway	2 314	5 695	10 033	21 294	52
Poland	n.a.	n.a.	1 356	2 615	21
Portugal	880	1 954	3 787	5 847	33
Romania	n.a.	n.a.	n.a.	1 640	direct
Russia	n.a.	n.a.	1 932	4 823	TNE
Slovakia	n.a.	n.a.	1 529	3 023	22
Slovenia	n.a.	n.a.	4 456	7 092	36
Spain	1 361	2 808	4 964	8 370	36
South Africa	n.a.	1 719	2 099	3 119	TNE
Sweden	2 798	5 822	8 576	13 663	46
Switzerland	4 934	10 124	14 701	21 156	64
Turkey	n.a.	n.a.	n.a.	1 283	direct
United Kingdom	2 190	4 638	7 896	13 294	46
United States	2 895	6 336	10 355	15 753	TNE
Mean	2 291	4 975	6 746	9 104	42

Notes. n.a. = not available, TNE = transnational estimator, direct = directly calculated in threshold table.

C Meta-Analyses and Reviews Pertinent in this Research Field

Anderson, L. M., Shinn, C., Fullilove, M. T., Scrimshaw, S. C., Fielding, J. E., Normand, J., ... & Task Force on Community Preventive Services. (2003). The effectiveness of early childhood development programs: A systematic review. *American Journal of Preventive Medicine*, *24*(3), 32–46. doi: 10.1016/S0749-3797(02)00655-4

Bakermans-Kranenburg, M. J., Van Ijzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin*, *129*(2), 195–215. doi: 10.1037/0033-2909.129.2.195

Barnett, W. S., & Boocock, S. S. (Eds.). (1998). *Early care and education for children in poverty: Promises, programs, and long-term results.* SUNY Press.

Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*, *25*(2), 140–165. doi: 10.1016/j.ecresq.2009.11.001

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *The Teachers College Record*, *112*(3), 579–620.

Chambers, B., Cheung, A., Slavin, R. E., Smith, D., & Laurenzano, M. (2010). *Effective early childhood education programs: A systematic review*. Center for Research and Reform in Education. Retrieved from ERIC database (ED527643).

Darrow, C. L. (2009). *Language and literacy effects of curriculum interventions for preschools serving economically disadvantaged children: A meta analysis*. Society for Research on Educational Effectiveness (SREE) Conference Abstract Template. Retrieved from ERIC database (ED514949).

Farrington, D. P., & Welsh, B. C. (2002). Family-based crime prevention. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp.

22–55). London: Routledge.

Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: A meta-analysis. *Australian & New Zealand Journal of Criminology*, *36*(2), 127–151. doi: 10.1375/acri.36.2.127

Gottfredson, D. C, Wilson, D. B, & Najaka, S. S. (2002): School-based crime prevention. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 55–164). London: Routledge.

Hahn, R., Fuqua-Whitley, D., Wethington, H., Lowy, J., Crosby, A., Fullilove, M., ... & Task Force on Community Preventive Services. (2007). Effectiveness of universal school-based programs to prevent violent and aggressive behavior: A systematic review. *American Journal of Preventive Medicine*, *33*(2), 114–129. doi: 10.1016/j.amepre.2007.04.012

Kim, J. S., & Quinn, D. M. (2013). The effects of summer reading on low-income children's literacy achievement from kindergarten to grade 8. A meta-analysis of classroom and home interventions. *Review of Educational Research*, *83*(3), 386–431. doi: 10.3102/0034654313483906

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*(2), 275–313.

Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *The Annals of the American Academy of Political and Social Science*, *587*(1), 84–109. doi: 10.1177/0002716202250793

MacLeod, J., & Nelson, G. (2000). Programs for the promotion of family wellness and the prevention of child maltreatment: A meta-analytic review. *Child Abuse & Neglect*, *24*(9), 1127–1149. doi: 10.1016/S0145-2134(00)00178-2

Manning, M., Homel, R., & Smith, C. (2010). A meta-analysis of the effects of early developmental prevention programs in at-risk populations on non-health outcomes in adolescence. *Children and Youth Services Review*, *32*(4), 506–519. doi: 10.1016/j.childyouth.2009.11.003

Nelson, G., Westhues, A., & MacLeod, J. (2003). A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment*, *6*(1), 31a. doi: 10.1037/1522-3736.6.1.631a

Piquero, A. R., Farrington, D. P., Welsh, B. C., Tremblay, R., & Jennings, W. G. (2009). Effects of early family/parent training programs on antisocial behavior and delinquency. *Journal of Experimental Criminology*, *5*(2), 83–120. doi: 10.1007/s11292-009-9072-x

Watt, N. F., Ayoub, C., Bradley, R. H., Puma, J. E., & LeBoeuf, W.A. (Eds.). (2006). *The crisis in youth mental health: Critical issues and effective programs, Vol. 4: Early intervention programs and policies.* Westport, C.T.: Praeger Publishers/Greenwood Publishing Group (Child psychology and mental health).

D Overview of the Applied Search Options by Database

On the subsequent page a table is depicted which contains information of the online search regarding the applied search options for the interfaces and the corresponding databases. Those include the used search fields (i.e., *Search fields*), proximity operators (i.e., *Proximity*), truncations (i.e., *T*), and wildcards (i.e., *WC*). Truncation are separated into open as well as restricted truncation (i.e., *o* and *r*, respectively) and wildcards are also separated into two types, namely optional (i.e., *opt*) and fixed (i.e., *f*). For detailed descriptions and applications of the search functions see appendix E.

Not listed in the table are the search option which all interfaces have in common, such as Boolean Operators, quotation marks and parentheses.

Interface	Database	Search fields	Proximity	T		WC	
				o	r	opt	f
OvidSP	MEDLINE	Title (X.ti.), Abstract (X.ab.), Author Keywords (X.kw.), Keyword Heading Words (X.kf.), MeSH (X.me.)	X adjΣ Y	*	#Σ	?	#
	ERIC	Titel (X.ti.), Abstract (X.ab.), ERIC Subject Headings (X.sh.)					
EBSCOhost	PsycINFO PSYNDEX PsycARTICLES	Title (TI X), Abstract (AB X), Keywords (KW X), Subjects (SU X), Age Groups (AG X), Population Codes (PO X), Classification Codes (CC X)	X wΣ Y	*		#	?
Thomson Reuters	Web of Science	Title (TI=X), Author Keywords (AK=X), Keywords Plus® (KP=X), Topic (TS=X), Research Area (SU=X), Web of Science Category (WC=X)	X near/Σ Y	*		$	?
Wiley Online Library	The Cochrane Library	Title (X:ti), Abstract (X:ab), Keywords (X:kw), MeSH (ohne Kürzel)	X near/Σ Y	*		?	

Continued on the next page...

Interface	Database	Search fields	Proximity	T		WC	
				o	r	opt	f
ProQuest	ASSIA, PAIS, ProQuest D&T, SSA, SA, WPSA	Title TI(X), Abstract AB(X), Subjects SU(X)	X n/Σ Y	*	$Σ		?

Notes. T = truncation, o = open truncation, r = restricted truncation, WC = wildcard, opt = optional wildcard, f = fixed wildcard, MeSH = Medical Subject Heading, X or Y = search term X or Y, Σ = number, ASSIA = Applied Social Sciences Index and Abstracts, PAIS = PAIS International, ProQuest D&T = ProQuest Dissertations & Theses Full Text, SSA = Social Services Abstracts, SA = Sociological Abstracts, WPSA = Worldwide Political Science Abstracts, AE/BE = American English/British English.

E Applied Search Options

Central search fields as well as replacement options will be defined and their application explained in this appendix.

Search Fields

Keywords. *Keywords* are search terms describe by the study author(s) to specify the content of an article. It is therefore a relatively unsystematic and subjective description system which allows an intuitive approach when developing a search strategy. In this quality, the keyword search corresponds to the search within the title or the abstract, where likewise information given by the authors build the basis for the search field. Important to mention is that this concerns the *search field* keyword and not the *term* keyword in general. The latter is simply used as a synonym for a central word, independently of the applied search field, and will be referred to as key term in the following to better distinguish the two.

Subjects. In contrast to keywords, *subjects* refers to pre-defined "controlled" vocabulary which is being provided by a database. Subjects (sometimes also referred to as descriptors) are assigned to the articles with regards to the content. This is usually done by a team composed of multidisciplinary professionals. Some databases provide hierarchical subjects, also referred to as thesaurus. The U.S. National Library of Medicine, for instance, offers the System of Medical Subject Heading (MeSH), which can be used in the database Medline. Thereby searches can be designed which find subjects simultaneously on different hierarchical levels. Prior to the integration of a subject in the search strategy, the effort to find the exact relevant descriptor needs to be made. However, using them usually results in finding a higher proportion of relevant articles, compared to the search via keywords.

Replacement Options

Quotation Marks and Parenthesis. Quotation marks enable *phrase searching* which means using a search term which consists of several words in that exact order (e.g., *economically disadvantaged*). This option can be applied in most of the search fields. In contrast to that, parentheses are needed to group and prioritize search terms. The expression *child AND poor OR poverty* yields articles which contain either the word

child and the word *poor* or articles which contain the word poverty. The prioritization in *child AND (poor OR poverty)* will only find articles that use the word child, along with the word poor or poverty. The interfaces chosen for the search all prioritise AND before OR or NOT, so that search *1 AND 2 NOT 3*, for instance, leads to the same results as *(1 AND 2) NOT 3*.

Proximity Operators. Proximity operators are a modification of Boolean operators. By their application, only articles will be found that includes all sought-after search terms within a certain distance to each other. The distance can be a fixed number of words or the co-occurrence of the search terms within one sentence or paragraph. The search command *poor adj2 population*, for example, inserted in the database Medline would only yield articles where the words *poor* and *population* appear within maximally two words in between (e.g., *poor rural population*, *population that is poor*, *urban poor population*). The advantage is the following: While *poor* in combination with *population* is a relatively good search command in order to identify poverty samples, the word *poor* alone, as a rather unspecific term with many different meanings, leads to an immoderate amount of search results.

Truncations. Yet another search refinement is possible with the truncation feature which is synonymous with scissoring or cutting something. Accordingly, by using the truncation symbol, an arbitrary number of letters is left open (open truncation) in order to search for variations of a word and not just for the word itself. A search for *adolescen**, performed in the database PsycINFO, for instance, results in articles in which the words *adolescent*, as well as *adolescents* or *adolescence* appear. Some databases offer the option to limit the number of released characters (restricted truncation). In the database Medline, for example, the search term *famil$3* yields articles containing the term *family* or *families*, leaving out the ones with the term *familiarity*. Truncations are especially useful for finding singular and plural forms as well as different word endings within a search.

Wildcards. An even more specific substitution of characters is possible by using wildcards. Wildcards are place holder signs that replace either zero or one character (optional wildcard) or precisely one character (fixed wildcard). For example in the database Web of Science, the optional wildcard *$* in the search term *neighbo$rhood* yields to both, articles that use the English spelling *neighbourhood* and the ones that

use the American spelling *neighborhood*. The fixed wildcard *?* in *wom?n* detects articles which contain the word *women*, as well as *woman*.

F Online Search Strategies

This appendix contains all applied online search strategies. The following table gives an overview of the five interfaces that the online search covered. The date refers to the date when the search has been conducted. For the update (in brackets) the same search strategy as the main search has been used only limited to the date of the last search until the end of 2013.

		Studies Found	
Interface	Date (Update)	w/ D	w/out D
OvidSP	04.09.2013 (25.11.2014)	2 721	2 500
EBSCOHost	30.01.2013 (28.01.2014)	1 349	896
Thomson Reuters	13.02.2013 (05.12.2014)	2 044	1 266
Wiley Online Library	06.01.2013 (18.12.2014)	941	381
ProQuest	02.04.2013 (28.02.2014)	1 324	986
Overall		8 379	6 029

Notes. w/ = with, w/out = without, D = duplicates.

Interface (Databases): OvidSP (MEDLINE, ERIC)

ID	Search terms
1	(counseling or intervention? or prevention? or program? or effectiveness).ti,ab,kw,kf. OR (exp counseling/ or exp therapy/ or exp intervention/ or exp prevention/ or exp programs/ or exp "educational research"/ or exp "educational methods"/ or exp "psychological studies"/ or exp improvement/ or exp achievement/ or exp "human services"/ or exp "psychology, applied"/ or exp "psychology, social"/ or exp "cost-benefit analysis"/) OR ("Program Effectiveness" or "Educational Experiments" or "poverty programs" or "Compensatory Education" or "Literacy Education").sh. OR ("Program Evaluation" or "Behavior Control" or "Social Work" or "Child Care" or "child health services" or "Early Intervention (Education) " or "Health Education").me.
2	("low* socioeconomic" or "low* ses" or poverty or "low* income" or "low* social class" or "economic hardship" or disadvantaged or (poor adj2 population?) or (poor adj1 communit$3) or (poor adj1 children) or (poor district?) or (poor adj1 area?) or (poor adj1 famil$3) or (poor adj2 wom#n) or (poor household?) or undereducated or (poor home environment?) or (poor adj1 neighbo?rhood?)).ti,ab,kw,kf. OR (exp "Poverty Areas"/ or exp Poverty/) OR ("Disadvantaged environment" or "Disadvantaged Youth" or "Economically Disadvantaged" or "Educational discrimination" or "Educationally Disadvantaged" or "low income" or "low income groups" or "Public Housing" or Underemployment or "Welfare Recipients").sh. OR (Unemployment or "Aid to families with dependent children").me.

Continued on the next page...

ID	Search terms
3	(Youth? or Teen* or Adolescen* or minors or Child* or Baby or Babies or Infant* or Kid? or Boy? or Girl?).ti,ab,kw,kf. OR (exp Children/ or exp Youth/ or exp "Elementary Secondary Education"/ or exp Child/ or exp Infant/ or exp Schools/) OR ("Early Adolescents" or "Adolescents" or "Late Adolescents" or Puberty or "Middle School Students" or "Elementary School Students" or "Junior High School Students" or "High School Freshmen" or "High School Seniors" or "Noncollege Bound Students" or "Grade 1" or "Grade 2" or "Grade 3" or "Grade 4" or "Grade 5" or "Grade 6" or "Grade 7" or "Grade 8" or "Grade 9" or "Grade 10" or "Grade 11" or "Grade 12" or "Intermediate Grades"or "Primary Education" or Kindergarten or "School Entrance Age").sh. OR (Adolescent or "Adolescent Psychology" or "Child Psychology" or Minors or students).me.
4	("control group?" or "controlled trial?" or "controlled study" or "comparison group").ti,ab,kw,kf. OR ("Control groups" or "experimental groups" or "matched groups").sh. OR ("Controlled Clinical Trial" or "Randomized Controlled Trial" or "Control Groups").me. OR "Controlled Clinical Trial as Topic"/ or "Randomized Controlled Trial as Topic"/
5	Limitation: Human only
6	(Exp "Chemical Actions and Uses"/ or exp "drug therapy"/ or exp "Complementary Therapies"/ or exp "inorganic chemicals"/ or exp "Complex Mixtures"/ or exp "Biological Factors"/) OR (Pharmacy or Pharmacology or "Money Management" or "Paying for College" or "Trusts (Financial)" or "Case studies").sh. OR ("Case-Control Studies" or "Cross-Sectional Studies").me
SEARCH = 1 AND 2 AND 3 AND 4 AND 5 NOT 6	

Interface (Databases): EBSCOhost (PsychINFO, PSYNDEX, PsycARTICLES)

ID	Search terms
S1	TI (Coaching or Counsel* or Effectiveness or Instruction* or Intervention* or Prevention* or Program* or Teaching or Therapy or Training# or Treatment or Tutoring) OR AB (Coaching or Counsel* or Effectiveness or Instruction* or Intervention* or Prevention* or Program* or Teaching or Therapy or Training# or Treatment or Tutoring) OR SU ("Active Listening" or "Anxiety Management" or Autohypnosis or Care or Caregiver# or "Change Strategies" or "Child Guidance" or "Classroom Management" or Coaching or "Cognitive Restructuring" or "Cognitive Techniques" or "Collaborative Learning" or "Community Mental Health Centers" or "Community Services" or "Computer Assisted Language Learning" or "Contingency Management" or "Cooperative Learning" or Cotherapy or Counsel* or Counterconditioning or "Debriefing (Psychological)" or "Discovery Teaching Method" or "Discussion Method" or Education or "Educational Field Trips" or "Educational Television" or Effectiveness* or "Experiential Learning" or "Family Preservation" or "Guided Imagery" or "Harm Reduction" or "Health Promotion" or "Home Schooling" or "Hot Line Services" or Hypnosis or Hypnotherapy or Instruction* or "Integrated Services" or Intervention or "Lecture Method" or "Mastery Learning" or Meditation or "Mental Health Services" or Mindfulness or "Montessori Method" or "Neurolinguistic Programming" or "Open Classroom Method" or Prayer or Prevention or "Problem Based Learning" or Program or "Project Follow Through" or "Project Head Start" or Psychoanalysis or Psychoeducation or "Psychotherapeutic Techniques" or Psychotherapy or "Public Health Services" or Relaxation* or "Religious Practices" or "Remedial Reading" or "Risk Management" or Scaffolding or "Self Help Techniques" or "Self Management" or "Self Regulated Learning" or "Social Group Work" or "Social Services" or Sociotherapy or "Stress Management" or "Student Personnel Services" or "Support Groups" or Teaching or Therapy or Training or Treatment or Tutoring or "Upward Bound" or "Virtual Classrooms" or Yoga)

Continued on the next page...

ID	Search terms
S2	TI ("economic hardship" or "economic* disadvantage*" or "financ* disadvantage*" or "low* income" or "low* ses" or "low* social class" or "low* socioeconomic" or "poor home environment*" or poverty or "socioeconomically depressed" or undereducated) OR AB ("economic hardship" or "economic* disadvantage*" or "financ* disadvantage*" or "low* income" or "low* ses" or "low* social class" or "low* socioeconomic" or "poor home environment*" or poverty or "socioeconomically depressed" or undereducated) OR poor w1 (children or district* or area* or family or families or communit* or household* or "home environment" or neighbo#rhood*) or poor w2 (population* or wom?n) OR SU ("Cultural Deprivation" or Disadvantaged or "Educational Financial Assistance" or "Financial Strain" or "Food Deprivation" or Ghettoes or Homeless or "Lower Class*" or "Lower Income Level" or Poverty* or "Social Deprivation" or "Stimulus Deprivation" or "Underinsured (Health Insurance)" or Unemployment or "Uninsured (Health Insurance)" or "Water Deprivation")
S3	TI (adolescen* or babies or baby or boy or child* or girl* or grader* or infant* or kid or minor or minors or pupil# or students or teen* or toddler* or youth*) OR AB (adolescen* or babies or baby or boy or child* or girl* or grader* or infant* or kid or minor or minors or pupil# or students or teen* or toddler* or youth*) OR KW (adolescen* or babies or baby or boy or child* or girl* or grader* or infant* or kid or minor or minors or pupil# or students or teen* or toddler* or youth*) OR AG (Childhood or Adolescence) OR TI Parent* OR SU ("Parent Training" or "Childrearing Practices" or *Parenting)

Continued on the next page...

ID	Search terms
S4	TI ("Control group*" or "Comparison group*" or "Controlled trial*" or "Comparison trial*" or "Control children" or "Comparison children" or "Control* stud*" or "Comparison stud*" or "Control famil*" or "Comparison famil*" or "Control school*" or "Comparison school*" or "Control class" or "Comparison class" or RCT or randomi* or "Treatment as usual") OR AB ("Control group*" or "Comparison group*" or "Controlled trial*" or "Comparison trial*" or "Control children" or "Comparison children" or "Control* stud*" or "Comparison stud*" or "Control famil*" or "Comparison famil*" or "Control school*" or "Comparison school*" or "Control class" or "Comparison class" or RCT or randomi* or "Treatment as usual") OR SU ("Control group*" or "Comparison group*" or "Controlled trial*" or "Comparison trial*" or "Control children" or "Comparison children" or "Control* stud*" or "Comparison stud*" or "Control famil*" or "Comparison famil*" or "Control school*" or "Comparison school*" or "Control class" or "Comparison class" or RCT or randomi* or "Treatment as usual" or "Between Groups Design" or "Experiment Controls")
S5	TI ("case-control stud*" or "cross-sectional stud*") OR CC (2500 or 2510 or 2520 or 2530 or 2540 or 2560 or 2580 or 2860) OR PO 20
SEARCH = S1 AND S2 AND S3 AND S4 NOT S5	

Interface (Database): Thomson Reuters (Web of Science)

ID	Search terms
#1	TS=("Active Listening" or "Anxiety Management" or Autohypnosis or "Child Guidance" or "Classroom Management" or Coaching or "Cognitive Restructuring" or "Cognitive Techniques" or "Collaborative Learning" or "Community Mental Health Centers" or "Community Services" or "Computer Assisted Language Learning" or "Contingency Management" or "Cooperative Learning" or Cotherapy or Counsel* or Counterconditioning or "Psychological Debriefing" or "Discovery Teaching Method" or "Discussion Method" or "Educational Field Trips" or "Educational Television" or "Experiential Learning" or "Family Preservation" or "Guided Imagery" or "Health Promotion" or "Home Schooling" or "Hot Line Services" or Hypnosis or Hypnotherapy or Instruction$ or "Integrated Services" or "Lecture Method" or "Mastery Learning" or Meditation or "Mental Health Services" or Mindfulness or "Montessori Method" or "Neurolinguistic Programming" or "Open Classroom Method" or "Problem Based Learning" or "Project Follow Through" or "Head Start" or Psychoanalysis or Psychoeducation or "Psychotherapeutic Techniques" or Psychotherapy or "Public Health Services" or Relaxation or "Remedial Reading" or "Risk Management" or "Self Help Techniques" or "Self Management" or "Self Regulated Learning" or "Social Group Work" or "Social Service" or "Stress Management" or "Student Personnel Services" or "Support Groups" or Teaching or Tutoring or "Upward Bound" or "Virtual Classrooms" or Yoga or ((effects or effectiveness or efficacy or efficiency or evaluat* or education*) near/5 (intervention$ or prevention$ or program$ or training$ or treatment$))) OR TI=(Training$ or Treatment$ or Program$ or Prevention$ or Intervention$) OR AK=(Training$ or Treatment$ or Program$ or Prevention$ or Intervention$) OR KP=(Training$ or Treatment$ or Program$ or Prevention$ or Intervention$)

Continued on the next page...

ID	Search terms
#2	TS=("Aid to families with dependent children" or "Cultural Deprivation" or Disadvantaged or "economic hardship" or "economic* disadvantage*" or "educational discrimination" or "Educational Financial Assistance" or "financ* disadvantage*" or "Financial Strain" or "Food Deprivation" or Ghettoes or "low* income" or "low* ses" or "low* social class" or "low* socioeconomic" or "Lower Class" or "Lower Income Level$" or "poor home environment*" or (poor near/1 (children or district* or area* or family or families or communit* or household* or "home environment" or neighbor$rhood*)) or (poor near/2 (population* or wom?n)) or poverty or "Public Housing" or "Social Deprivation" or "socioeconomically depressed" or "Stimulus Deprivation" or "Subsidi?ed Housing" or undereducated or Underinsured or Unemployment or Uninsured or "Welfare Recipients")
#3	TS=(adolescen* or babies or baby or boy$ or child* or "Childrearing Practices" or "Elementary Education" or "Elementary Secondary Education" or Girl$ or Grader$ or "High School Freshmen" or "High School Seniors" or Infancy or Infant$ or Kid$ or Kindergarten or Minor$ or Neonat* or "Parent Training" or Preadolescent$ or "Preschool Age" or "Primary Education" or Puberty or pupils or "School Age" or "School Entrance Age" or "Secondary Education" or Student$ or Teen$ or Toddler$ or Youth$)
#4	TS=(Control* near/2 (group$ or trial$ or children or stud* or famil* or school$ or class or condition$) or Comparison near/2 (group$ or trial$ or children or stud* or famil* or school$ or class or condition$) or "Between Group Design" or RCT$ or (Random* near/2 control*) or "Treatment as usual" or "Experimental groups")

Continued on the next page...

ID	Search terms
#5	TI=("case-control stud*" or "cross-sectional stud*") OR SU=(Agriculture or Anesthesiology or "Biochemistry & Molecular Biology" or "Biodiversity & Conservation" or Biophysics or "Biotechnology & Applied Microbiology" or "Cell Biology or Entomology" or "Evolutionary Biology" or Fisheries or "Food Science & Technology" or Forestry or "Genetics & Heredity" or "Geriatrics & Gerontology" or "Marine & Freshwater Biology" or "Mathematical & Computational Biology" or "Medical Ethics" or "Medical Informatics" or "Medical Laboratory Technology" or Microbiology or Mycology or "Neurosciences & Neurology" or "Pharmacology & Pharmacy" or "Plant Sciences" or "Radiology, Nuclear Medicine & Medical Imaging" or Rheumatology or Surgery or Transplantation or "Veterinary Sciences" or Zoology or "Astronomy & Astrophysics" or Chemistry or Crystallography or Electrochemistry or "Geochemistry & Geophysics" or Geology or Mathematics or "Meteorology & Atmospheric Sciences" or Mineralogy or "Mining & Mineral Processing" or Oceanography or Optics or "Physical Geography" or Physics or "Polymer Science" or Thermodynamics or "Water Resources" or "Automation & Control Systems" or "Construction & Building Technology" or "Energy & Fuels" or Engineering or "Materials Science" or Mechanics or "Metallurgy & Metallurgical Engineering" or Microscopy or "Operations Research & Management Science" or "Remote Sensing" or Robotics or "Science & Technology Other Topics" or Spectroscopy or Telecommunications or Transportation or Architecture or Archaeology or Geography) OR WC=(Agronomy or "Biochemical Research Methods" or Biology or "Chemistry, Analytical" or "Chemistry, Applied" or "Chemistry, Inorganic & Nuclear" or "Chemistry, Medicinal" or "Chemistry, Multidisciplinary" or "Chemistry, Organic" or "Chemistry, Physical" or "Clinical Neurology" or Ergonomics or "Industrial Relations & Labor" or Law or Limnology or "Medicine, Legal" or "Medieval & Renaissance Studies" or "Nanoscience & Nanotechnology" or Neuroimaging or Ornithology or "Soil Science")
SEARCH = #1 AND #2 AND #3 AND #4 NOT #5	

Interface (Database): Wiley Online Library (The Cochrane Library)

The search strategy for Cochrane differs from the previous ones, because MeSH Descriptors can only be searched for separately. In order to retain an overview, the blocks are assigned to the equivalent ID number of the search part in brackets.

ID	Search terms
#1	(coaching or counseling or effectiveness or efficacy or efficiency or instruction? or intervention? or prevention? or program? or teaching or therapy or training? or treatment or tutoring):ti,ab,kw OR ("Active Listening" or "Anxiety Management" or Autohypnosis or "Child Guidance" or "Classroom Management" or "Cognitive Restructuring" or "Cognitive Techniques" or "Collaborative Learning" or "Community Mental Health Centers" or "Community Services" or "Computer Assisted Language Learning" or "Contingency Management" or "Cooperative Learning" or Cotherapy or Counterconditioning or "Discussion Method" or "Educational Field Trips" or "Educational Television" or "Experiential Learning" or "Family Preservation" or "Guided Imagery" or "Health Promotion" or "Home Schooling" or "Hot Line Services" or Hypnosis or Hypnotherapy or "Integrated Services" or "Lecture Method" or "Mastery Learning" or Meditation or "Mental Health Services" or Mindfulness or "Montessori Method" or "Neurolinguistic Programming" or "Open Classroom Method" or "Problem Based Learning" or "Project Follow Through" or "Head Start" or Psychoanalysis or Psychoeducation or "Psychotherapeutic Techniques" or Psychotherapy or "Public Health Services" or Relaxation or "Remedial Reading" or "Risk Management" or "Self Help Techniques" or "Self Management" or "Self Regulated Learning" or "Social Group Work" or "Social Service" or "Stress Management" or "Student Personnel Services" or "Support Groups" or "Upward Bound" or "Virtual Classrooms" or Yoga):ti,ab,kw (Word variations have been searched)
#2	MeSH descriptor: [Counseling] explode all trees
#3	MeSH descriptor: [Early Intervention (Education)] explode all trees
#4	MeSH descriptor: [Teaching] explode all trees

Continued on the next page...

ID	Search terms
#5	MeSH descriptor: [Program Evaluation] explode all trees
#6	MeSH descriptor: [Child Care] explode all trees
#7	#1 OR #2 OR #3 OR #4 OR #5 OR #6 (**Block 1**)
#8	("low* socioeconomic" or "low* ses" or "low* income" or "low* social class" or poverty or "economic hardship" or "economic* disadvantage*" or "financ* disadvantage*" or undereducated or "socioeconomically depressed" or Underinsured or Unemployment or Uninsured or "disadvantaged environment" or "disadvantaged Youth" or "Educational discrimination" or "Educationally Disadvantaged" or Underemployment or "Welfare Recipients" or "Public Housing" or "Subsidi?ed Housing" or "Cultural Deprivation" or "educational discrimination" or "Educational Financial Assistance" or "Financial Strain" or "Food Deprivation" or Ghettoes or "Social Deprivation" or "Stimulus Deprivation") (Word variations have been searched; All Text Search)
#9	(poor near/2 (children or district? or area? or famil* or communit* or household? or "home environment" or neighbor?rhood? or population* or wom?n)) (Word variations have been searched; All Text Search)
#10	MeSH descriptor: [Poverty Areas] explode all trees
#11	MeSH descriptor: [Poverty] explode all trees
#12	MeSH descriptor: [Unemployment] explode all trees
#13	MeSH descriptor: [Public Assistance] explode all trees
#14	MeSH descriptor: [Aid to Families with Dependent Children] explode all trees
#15	#8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 (**Block 2**)

Continued on the next page...

ID	Search terms
#16	(adolescen* or babies or baby or boy? or child* or Girl? or Grader? or Infancy or Infant? or Kid? or Kindergarten or Minor? or Neonat* or Preadolescent? or Student? or Teen? or Toddler? or Youth? or "Early Adolescents" or "Adolescents" or "Late Adolescents" or Puberty or "Middle School Students" or "Elementary School Students" or "Junior High School Students" or "High School Freshmen" or "High School Seniors" or "Noncollege Bound Students" or "Intermediate Grades" or "Primary Education" or Kindergarten or "School Entrance Age" or "Childrearing Practices" or "Elementary Education" or "Elementary Secondary Education" or "Parent Training" or "Preschool Age" or "Primary Education" or pupils or "School Age" or "Secondary Education" or "Childrearing Practices" or *Parenting):ti,ab,kw (Word variations have been searched)
#17	MeSH descriptor: [Child] explode all trees
#18	MeSH descriptor: [Infant] explode all trees
#19	MeSH descriptor: [Adolescent] explode all trees
#20	MeSH descriptor: [Schools] explode all trees
#21	MeSH descriptor: [Adolescent Psychology] explode all trees
#22	MeSH descriptor: [Child Psychology] explode all trees
#23	MeSH descriptor: [Students] explode all trees
#24	MeSH descriptor: [Minors] explode all trees
#25	#16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 (**Block 3**)
#26	((control* or comparison) near/2 (group? or trial? or children or stud* or famil* or school? or class or condition?)):ti,ab,kw OR (Random* near/2 control*):ti,ab,kw (Word variations have been searched)
#27	(RCT or "Treatment as usual" or "Between Groups Design" or "Experiment Controls"):ti,ab,kw (Word variations have been searched)
#28	MeSH descriptor: [Randomized Controlled Trial] explode all trees
#29	MeSH descriptor: [Control Groups] explode all trees

Continued on the next page...

ID	Search terms
#30	MeSH descriptor: [Controlled Clinical Trials as Topic] explode all trees
#31	MeSH descriptor: [Randomized Controlled Trials as Topic] explode all trees
#32	#26 OR #27 OR #28 OR #29 OR #30 OR #31 (**Block 4**)
#33	MeSH descriptor: [Chemical Actions and Uses] explode all trees
#34	MeSH descriptor: [Drug Therapy] explode all trees
#35	MeSH descriptor: [Complementary Therapies] explode all trees
#36	MeSH descriptor: [Inorganic Chemicals] explode all trees
#37	MeSH descriptor: [Complex Mixtures] explode all trees
#38	MeSH descriptor: [Biological Factors] explode all trees
#39	MeSH descriptor: [Cross-Sectional Studies] explode all trees
#40	MeSH descriptor: [Case-Control Studies] explode all trees
#41	#33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 (**Block 5**)
#42	#7 AND #15 AND #25 AND #32 NOT #41 (**SEARCH**)

Interface (Databases): ProQuest (Applied Social Sciences Index and Abstracts (ASSIA), PAIS International (PAIS), ProQuest Dissertations & Theses Full Text (ProQuest D&T), Social Services Abstracts (SSA), Sociological Abstracts (SA), Worldwide Political Science Abstracts (WPSA))

The search strategy for the interface ProQuest had to be run separately for some databases because the thesauri of one database caused search errors for others. That was the case for Applied Social Sciences Index and Abstracts, as well as for Sociological Abstract and for Social Services Abstracts. Hence, three slightly adapted strategies for the different databases are given below.

Databases: PAIS , ProQuest D&T, WPSA

ID	Search terms
1	TI,AB(counseling or intervention or prevention or teaching or treatment or tutoring) OR SU("After School Programs" or "Boys education" or "Child care" or "Cognitive behavioural counseling" or "Compensatory education" or "Computer assisted counseling" or "Computer assisted instruction" or "Crime prevention" or "Day care" or "Early childhood education" or "Early intervention programmes" or "Educational guidance" or "Elementary education" or "Girls education" or "Group counseling" or "Head Start" or "Health Care Services" or "Humanistic counseling" or "Instructions" or "Interactive computer assisted learning" or "Long term counseling" or "Meditation" or "Montessori method" or "Multicultural counseling" or "Out of school care" or "Outplacement services" or "Parenthood education" or "Peer group counseling" or "Prevention" or "Psychodynamic counseling" or "Public health education" or "Reading instruction" or "School counseling" or "School music programs" or "Secondary education" or "Selfcounselling" or "Sex education" or "Summer schools" or "Teaching" or "Tutorials" or "Tutoring" or "Yoga")

Continued on the next page...

ID	Search terms
2	TI,AB("low* socioeconomic" or "low* ses" or "low* income" or "low* social class" or poverty or "economic hardship" or "economic* disadvantage*" or "financ* disadvantage*" or undereducated or "socioeconomically depressed" or "disadvantaged environment" or "disadvantaged Youth" or "Educational discrimination" or "Educationally Disadvantaged" or "Educational Financial Assistance" or "Financial Strain") OR TI,AB(poor near/1 (children or district or area or family or community or household or "home environment" or neighborhood or people) or (poor near/2 (population or wom?n))) OR SU("At risk youth" or "At risk students" or "*Poverty" or "Economically depressed areas" or "Ghetto*s" or "Housing Subsidies" or "Low Income*" or "Low status people" or "Lower Class" or "Public Housing" or "Socially handicapped" or "Underclass")
3	TI,AB(adolescen* or baby or boy or child* or Girl or Grader or Infant or Kid or Minor or Preadolescent or Student or Teen* or Toddler or Youth or pupil) OR SU("Adolescents" or "Babies" or "Boys" or "Child rearing" or "Childrearing Practices" or "Children" or "Elementary School Students" or "Girls" or "High School Students" or "Infants" or "Junior High School Students" or "Newborn babies" or "Parent management training" or "Parent Training" or "Parenting" or "Preschool Children" or "Pupils" or "Students" or "Young children" or "Young people" or "Youth")
4	TI,AB(((control* or comparison) near/2 (group or trial or children or study or family or school or class or condition)) or (Random* near/2 control*) or RCT or "Treatment as usual" or "Between Groups Design" or "Experiment Controls") OR SU("Control groups" or "Randomized controlled trials")
SEARCH = 1 AND 2 AND 3 AND 4	

Databases: ASSIA/SA

ID	Search terms
1	TI,AB(counseling or intervention or prevention or teaching or treatment or tutoring)
2	TI,AB("low* socioeconomic" or "low* ses" or "low* income" or "low* social class" or poverty or "economic hardship" or "economic* disadvantage*" or "financ* disadvantage*" or undereducated or "socioeconomically depressed" or "disadvantaged environment" or "disadvantaged Youth" or "Educational discrimination" or "Educationally Disadvantaged" or "Educational Financial Assistance" or "Financial Strain") OR TI,AB(poor near/1 (children or district or area or family or community or household or "home environment" or neighborhood or people) or (poor near/2 (population or wom?n))) OR SU("*Poverty" or "Economically depressed areas" or "Ghetto*s" or "Housing Subsidies" or "Low Income*")
3	TI,AB(adolescen* or baby or boy or child* or Girl or Grader or Infant or Kid or Minor or Preadolescent or Student or Teen* or Toddler or Youth or pupil) OR SU("Adolescents" or "Babies" or "Boys" or "Child rearing" or "Children" or "Elementary School Students" or "Girls" or "Parent management training" or "Parenting" or "Preschool Children" or "Pupils" or "Students" or "Young children" or "Young people")
4	TI,AB(((control* or comparison) near/2 (group or trial or children or study or family or school or class or condition)) OR (Random* near/2 control*) OR RCT OR "Treatment as usual") OR SU("Control groups" OR "Randomized controlled trials")
	SEARCH = 1 AND 2 AND 3 AND 4

Databases: SSA

ID	Search terms
1	TI,AB(counseling or intervention or prevention or teaching or treatment or tutoring)
2	TI,AB("low* socioeconomic" or "low* ses" or "low* income" or "low* social class" or poverty or "economic hardship" or "economic* disadvantage*" or "financ* disadvantage*" or undereducated or "socioeconomically depressed" or "disadvantaged environment" or "disadvantaged Youth" or "Educational discrimination" or "Educationally Disadvantaged" or "Educational Financial Assistance" or "Financial Strain") OR TI,AB(poor near/1 (children or district or area or family or community or household or "home environment" or neighborhood or people) or (poor near/2 (population or wom?n))) OR SU("*Poverty" or "Ghetto*s" or "Public Housing" or "Low Income*" or "Lower Class" or "Underclass")
3	TI,AB(adolescen* or baby or boy or child* or Girl or Grader or Infant or Kid or Minor or Preadolescent or Student or Teen* or Toddler or Youth or pupil) OR SU("Adolescents" or "Childrearing Practices" or "Children" or "Elementary School Students" or "High School Students" or "Infants" or "Junior High School Students" or "Parent Training" or "Preschool Children" or "Youth")
4	TI,AB(((control* or comparison) near/2 (group or trial or children or study or family or school or class or condition)) or (Random* near/2 control*) or RCT or "Treatment as usual" or "Between Groups Design" or "Experiment Controls") OR SU("Control groups" or "Randomized controlled trials")
SEARCH = 1 AND 2 AND 3 AND 4	

G Included Primary Studies

On the following pages the included primary studies are listed. The main studies are unmarked and the corresponding duplicates are inserted directly below the main study. Hereby, supplemental duplicates are marked with a black circle (•), whereas partial or complete duplicates are preceded with a white circle (◦). For more information regarding duplicates, see section 2.1.6.

Abbott-Shim, M., Lambert, R. & McCarty, F. (2003). A comparison of school readiness outcomes for children randomly assigned to a head start program and the program's wait list. *Journal of Education for Students Placed at Risk*, *8*(2), 191–214. doi: 10.1207/S15327671ESPR0802_2

Allor, J. & McCathren, R. (2004). The efficacy of an early literacy tutoring program implemented by college students. *Learning Disabilities Research & Practice*, *19*(2), 116–129. doi: 10.1111/j.1540-5826.2004.00095.x

Andrews, S. R., Blumenthal, J. B., Johnson, D. L., Kahn, A. J., Ferguson, C. J., Lasater, T. M., ... & Wallace, D. B. (1982). The skills of mothering: A study of parent child development centers. *Monographs of the Society for Research in Child Development*, *47*(6), 1–83. doi: 10.2307/1165919

- Bridgeman, B., Blumenthal, J. B. & Andrews, S. R. (1981). *Parent Child Development Center: Final evaluation report.* Washington, DC: Department of Health and Human Services. Office of Human Development Services. Retrieved from ERIC database (ED213764).

- Johnson, D. L. (2006). Parent-Child Development Center follow-up project: Child behavior problem results. *Journal of Primary Prevention*, *27*(4), 391–407. doi: 10.1007/s10935-006-0040-1

- Johnson, D. L. & Blumenthal, J. (2004). The Parent Child Development Centers and school achievement: A follow-up. *Journal of Primary Prevention*, *25*(2), 195–209. doi: 10.1023/B:JOPP.0000042390.33338.91

- Johnson, D. L. & Breckenridge, J. N. (1982). The Houston parent—child development center and the primary prevention of behavior problems in young children. *American*

Journal of Community Psychology, *10*(3), 305–316. doi: 10.1007/BF00896497

- Johnson, D. L. & Walker, T. (1987). Primary prevention of behavior problems in Mexican-American children. *American Journal of Community Psychology*, *15*(4), 375–385. doi: 10.1007/BF00915208

- Walker, T. & Johnson, D. L. (1988). A follow-up evaluation of the Houston Parent-Child Development Center: Intelligence test results. *The Journal of Genetic Psychology*, *149*(3), 377–381. doi: 10.1080/00221325.1988.10532165

◦ Bridgeman, B., Hilton, T., Blumenthal, J. & Andrews, S. (1977). *Parent Child Development Centers: Long-term and short-term results.* Princeton, NJ: Educational Testing Service. Retrieved from ERIC database (ED195327).

◦ Johnson, D. L. (1976). *Parent Education and the Educationally Disadvantaged Child.* Paper presented at the 84th annual meeting of the American Psychological Association, Washington, D. C. Retrieved from ERIC database (ED127355).

◦ Johnson, D. L. (1981). The influence of an intensive parent education program on behavioral continuity of mothers and children. *Child Study Journal*, *11*(4), 187–199.

◦ Johnson, D. L. & Breckenridge, J. N. (1981). Effect of task on mother-child interaction results. Journal of Behavioral Assessment, *3*(4), 343–354. doi: 10.1007/BF01350837

◦ Johnson, D. L. & Walker, T. (1991). A follow-up evaluation of the Houston Parent-Child Development Center: School performance. *Journal of Early Intervention*, *15*(3), 226–236. doi: 10.1177/105381519101500302

◦ Johnson, D. L., Kahn, A. J. & Leler, H. (1976). *Houston Parent-Child Development Center, Final Report.* Houston, TX: Parent Child Development Center. Retrieved from ERIC database (ED135459).

◦ Johnson, D. L., Kahn, A. J., Hines, R. P., Leler, H. & Torres, M. (1976). *Measuring the learning environment of Mexican-American families in a parent education program.* Houston, TX: Parent Child Development Center: Retrieved from ERIC database (ED127355).

◦ Johnson, D. L., Leler, H., Rios, L., Brandt, L., Kahn, A. J., Mazeika, E., ... & Bisett, B.

(1974). The Houston Parent-Child Development Center: A parent education program for Mexican-American families. *American Journal of Orthopsychiatry*, *44*(1), 121–128. doi: 10.1111/j.1939-0025.1974.tb00876.x

Apel, K. & Diehm, E. (2013). Morphological awareness intervention with kindergarteners and first and second grade students from low SES homes: A small efficacy study. *Journal of Learning Disabilities*, *47*(1), 65–75. doi: 10.1177/0022219413509964

Aram, D., Fine, Y. & Ziv, M. (2013). Enhancing parent–child shared book reading interactions: Promoting references to the book's plot and socio-cognitive themes. *Early Childhood Research Quarterly*, *28*(1), 111–122. doi: 10.1016/j.ecresq.2012.03.005

Austin, D. B. (2005). The effects of a strengths development intervention program upon the self-perceptions of students' academic abilities (Doctoral dissertation). *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, *66*(05), 1631A.

Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R. & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, *23*(3), 299–313. doi: 10.1016/j.ecresq.2008.03.001

- Diamond, A., Barnett, W. S., Thomas, J. & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855) , 1387–1388. doi: 10.1126/science.1151148

Beckwith, L. (1988). Intervention with disadvantaged parents of sick preterm infants. *Psychiatry*, *51*(3), 242–247. doi: 10.1521/00332747.1988.11024398

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., ... & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, *79*(6), 1802–1817. doi: 10.1111/j.1467-8624.2008.01227.x

- Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C. & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology*, *20*(03), 821–843. doi: 10.1017/S0954579408000394

- Bierman, K. L., Nix, R. L., Heinrichs, B. S., Domitrovich, C. E., Gest, S. D., Welsh, J. A. &

Gill, S. (2014). Effects of Head Start REDI on children's outcomes 1 year later in different kindergarten contexts. *Child Development*, *85*(1), 140–159. doi: 10.1111/cdev.12117

- Nix, R. L., Bierman, K. L., Domitrovich, C. E. & Gill, S. (2013). Promoting children's social-emotional skills in preschool can enhance academic and behavioral functioning in kindergarten: Findings from Head Start REDI. *Early Education & Development*, *24*(7), 1000–1019. doi: 10.1080/10409289.2013.825565

Black, M. M., Dubowitz, H., Hutcheson, J., Berenson-Howard, J. & Starr, R. H. (1995). A randomized clinical trial of home intervention for children with failure to thrive. *Pediatrics*, *95*(6), 807–814.

- Black, M. M., Dubowitz, H., Krishnakumar, A. & Starr, R. H. (2007). Early intervention and recovery among children with failure to thrive: follow-up at age 8. *Pediatrics*, *120*(1), 59–69. doi: 10.1542/peds.2006-1657

- Hutcheson, J. J., Black, M. M., Talley, M., Dubowitz, H., Howard, J. B., Star, R. H. & Thompson, B. S. (1997). Risk status and home intervention among children with failure-to-thrive: Follow-up at age 4. *Journal of Pediatric Psychology*, *22*(5), 651–668. doi: 10.1093/jpepsy/22.5.651

Bleeker, M., James-Burdumy, S., Beyler, N., Dodd, A. H., London, R. A., Westrich, L., ... & Castrechini, S. (2012). *Findings from a randomized experiment of Playworks: Selected results from cohort 1.* Mathematica Policy Research, Inc. Retrieved from ERIC database (ED531230).

Borman, G. D. & Dowling, N. M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial. *Educational Evaluation and Policy Analysis*, *28*(1), 25–48. doi: 10.3102/01623737028001025

Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A. & Chambers, B. (2005a). Success for All: First-year results from the national randomized field trial. *Educational Evaluation and Policy Analysis*, *27*(1), 1–22. doi: 10.3102/01623737027001001

- Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A. & Chambers, B. (2005b). The national randomized field trial of Success for All: Second-year outcomes. *American Educational Research Journal*, *42*(4), 673–696. doi:

10.3102/00028312042004673

- Borman, G. D., Slavin, R. E., Cheung, A. C., Chamberlain, A. M., Madden, N. A. & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. American *Educational Research Journal*, *44*(3), 701–731. doi: 10.3102/0002831207306743

Boyce, L. K., Innocenti, M. S., Roggman, L. A., Norman, V. K. J. & Ortiz, E. (2010). Telling stories and making books: Evidence for an intervention to help parents in migrant Head Start families support their children's language and literacy. *Early Education and Development*, *21*(3), 343–371. doi: 10.1080/10409281003631142

Cappella, E., Hamre, B. K., Kim, H. Y., Henry, D. B., Frazier, S. L., Atkins, M. S. & Schoenwald, S. K. (2012). Teacher consultation and coaching within mental health practice: classroom and child effects in urban elementary schools. *Journal of Consulting and Clinical Psychology*, *80*(4), 597–610. doi: 10.1037/a0027725

Carter, J. L. (1967). *The long range effects of a language stimulation program upon Negro educationally disadvantaged first grade children.* University of Houston, TX. Retrieved from ERIC database (ED013276).

- Carter, J. L. (1969). *Follow-up Study of the effects of a language stimulation program upon Negro educationally disadvantaged first grade children. Final report.* University of Houston, TX. Retrieved from ERIC database (ED035714).

Chaya, M. S., Nagendra, H., Selvam, S., Kurpad, A. & Srinivasan, K. (2012). Effect of yoga on cognitive abilities in schoolchildren from a socioeconomically disadvantaged background: a randomized controlled study. *The Journal of Alternative and Complementary Medicine*, *18*(12), 1161–1167. doi: 10.1089/acm.2011.0579

Cirino, P. T., Vaughn, S., Linan-Thompson, S., Cardenas-Hagan, E., Fletcher, J. M. & Francis, D. J. (2009). One-year follow-up outcomes of Spanish and English interventions for English language learners at risk for reading problems. *American Educational Research Journal*, *46*(3), 744–781. doi: 10.3102/0002831208330214

- Vaughn, S., Cirino, P. T., Linan-Thompson, S., Mathes, P. G., Carlson, C. D., Hagan, E.

C., ... & Francis, D. J. (2006). Effectiveness of a Spanish intervention and an English intervention for English-language learners at risk for reading problems. *American Educational Research Journal*, *43*(3), 449–487. doi: 10.3102/00028312043003449

- Vaughn, S., Cirino, P. T., Tolar, T., Fletcher, J. M., Cardenas-Hagan, E., Carlson, C. D. & Francis, D. J. (2008). Long-term follow-up of Spanish and English interventions for first-grade English language learners at risk for reading problems. *Journal of Research on Educational Effectiveness*, *1*(3), 179–214. doi: 10.1080/19345740802114749

- Vaughn, S., Linan-Thompson, S., Mathes, P. G., Cirino, P. T., Carlson, C. D., Pollard-Durodola, S. D., ... & Francis, D. J. (2006). Effectiveness of Spanish intervention for first-grade English language learners at risk for reading difficulties. *Journal of Learning Disabilities*, *39*(1), 56–73. doi: 10.1177/00222194060390010601

  - Linan-Thompson, S., Vaughn, S., Prater, K. & Cirino, P. T. (2006). The response to intervention of English language learners at risk for reading problems. *Journal of Learning Disabilities*, *39*(5), 390–398. doi: 10.1177/00222194060390050201

  - Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., ... & Francis, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal*, *107*(2), 153–180. doi: 10.1086/510653

Cronan, T. A., Brooks, L. B., Kilpatrick, K., Bigatti, S. M. & Tally, S. (1999). The effects of a community-based literacy program: One-year follow-up findings. *Journal of Community Psychology*, *27*(4), 431–442. doi: 10.1002/(SICI)1520-6629(199907)27:4<431::AID-JCOP5>3.0.CO;2-V

Cronan, T. A., Cruz, S. G., Arriaga, R. I. & Sarkin, A. J. (1996). The effects of a community-based literacy program on young children's language and conceptual development. *American Journal of Community Psychology*, *24*(2), 251–272. doi: 10.1007/BF02510401

Curry, E. S. (1979). *The retention of specific language skills, acquired through a short-term, intensive teacher-directed program, among economically disadvantaged preschoolers* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 8100730).

Darrah, P. E. (1986). *The effects of interpersonal skills training on academic, behavioral, and person-*

ality characteristics of low-income high school students (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 8611831).

Davis, L. A. J. (1994). *The effect of parent involvement training on the achievement of Hispanic students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 9424368).

Deloria, D., Love, J .M., Goedinghaus, L., Gordon, S., Hanvey, R., Hockman, E., ... & Wilbur, J. (1974). *The National Home Start evaluation. Interim report V: Summative evaluation results.* Cambridge, MA: High/Scope Educational Research Foundation; Ypsilanti, MI: Abt Associates Inc. Retrieved from ERIC database (ED107380).

- Goodrich, N., Nauta, M., Rubin, A., Stein, M., Deloria, D. & Ruopp, R. (1975). *National Home Start evaluation interim report VI. Twelve-month program issues, outcomes and costs.* Cambridge, MA: High/Scope Educational Research Foundation; Ypsilanti, MI: Abt Associates Inc. Retrieved from ERIC database (ED134322).

- Love, J. M., Nauta, M., Coelen, C. G., Ruopp, R. & Deloria, D. (1975). *National Home Start evaluation interim report VI. Executive summary: Findings and recommendations.* Cambridge MA: High/Scope Educational Research Foundation, Ypsilanti; MI: Abt Associates Inc. Retrieved from ERIC database (ED134321).

- Love, J. M., Nauta, M. J., Coelen, C. G., Hewett, K. & Ruopp, R. R. (1976). *National Home Start evaluation: Final report. Findings and implications.* Cambridge, MA: High/Scope Educational Research Foundation; Ypsilanti, MI: Abt Associates Inc. Retrieved from ERIC database (ED134314).

- Deloria, D., Love, J .M., Goedinghaus, L., Gordon, S., Hanvey, R., Hockman, E., ... & Morris, K. (1974). *The National Home Start Evaluation. Interim Report IV: Summative evaluation results.* Cambridge, MA: High/Scope Educational Research Foundation; Ypsilanti, MI: Abt Associates Inc.

Denton, C. A., Solari, E. J., Ciancio, D. J., Hecht, S. A. & Swank, P. R. (2010). A pilot study of a kindergarten summer school reading program in high-poverty urban schools. *The Elementary School Journal*, *110*(4), 423–439. doi: 10.1086/651190

Diaz, I. (2010). *The Effect of morphological instruction in improving the spelling, vocabulary, and*

reading comprehension of high school english language learners (ELLs) (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (ProQuest Document ID 854554167).

Di Lorenzo, L. T. & Salter, R. (1967). *Second-year report on an evaluative study of prekindergarten programs for educationally disadvantaged children.* Paper presented at the 1967 Annual Convocation of the Educational Research Association, New York State. Retrieved from ERIC database (ED016523).

- Di Lorenzo, L. T. (1969). *Prekindergarten programs for educationally disadvantaged children. Final report.* New York State Education Department, Albany. Retrieved from ERIC database (ED038460).

Dixon, P., Schagen, I. & Seedhouse, P. (2011). The impact of an intervention on children's reading and spelling ability in low-income schools in India. *School Effectiveness and School Improvement*, *22*(4), 461–482. doi: 10.1080/09243453.2011.625125

Doherty, R. W. & Hilberg, R. S. (2008). Efficacy of five standards in raising student achievement: Findings from three studies. *The Journal of Educational Research*, *101*(4), 195–206. doi: 10.3200/JOER.101.4.195-206

Dwyer, R. C. (1972). *An evaluation of the effectiveness of a new type of preschool compensatory program: Environmental Academics.* Paper presented at the 1972 annual meeting of the American Educational Research Association, Chicago, IL. Retrieved from ERIC database (ED061273).

Dyson, N. I. (2011). *A number sense intervention for urban kindergartners at risk for mathematics difficulties* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 3465744).

- Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilitie*, *46*(2), 166–181. doi: 10.1177/0022219411410233

Edmonds, E., O'Donoghue, C., Spano, S. & Algozzine, R. F. (2009). Learning when school is out. *The Journal of Educational Research*, *102*(3), 213–222. doi: 10.3200/JOER.102.3.213-

222

Farr, B. (2006). *Report of a field study of quality teaching for English learners (QTEL).* San Francisco, CA: WestEd. Retrieved from rockman.com/projects/127.wested.qtel/QTEL_final_report.pdf

Farver, J. A. M., Lonigan, C. J. & Eppe, S. (2009). Effective early literacy skill development for young Spanish-speaking English language learners: An Experimental Study of Two Methods. *Child Development*, *80*(3), 703–719. doi: 10.1111/j.1467-8624.2009.01292.x

- Goodrich, J. M., Lonigan, C. J. & Farver, J. M. (2013). Do early literacy skills in children's first language promote development of skills in their second language? An experimental evaluation of transfer. *Journal of Educational Psychology*, *105*(2), 414–426. doi: 10.1037/a0031780

Fischel, J. E., Bracken, S. S., Fuchs-Eisenberg, A., Spira, E. G., Katz, S. & Shaller, G. (2007). Evaluation of curricular approaches to enhance preschool early literacy skills. *Journal of Literacy Research*, *39*(4), 471–501. doi: 10.1080/10862960701675333

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., ... & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement.* NCEE 2008-4030. National Center for Education Evaluation and Regional Assistance. Retrieved from http://www.mdrc.org/publication/impact-two-professional-development-interventions-early-reading-instruction-and/file-full

Ginsburg-Block, M. D. & Fantuzzo, J. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology*, *90*(3), 1–10. doi: 10.1037/0022-0663.90.3.560

Gonzales, N. A., Dumka, L. E., Millsap, R. E., Gottschall, A., McClain, D. B., Wong, J. J., ... & Kim, S. Y. (2012). Randomized trial of a broad preventive intervention for Mexican American adolescents. *Journal of Consulting and Clinical Psychology*, *80*(1), 1–16. doi: 10.1037/a0025063

- Carpentier, F. R. D., Mauricio, A. M., Gonzales, N. A., Millsap, R. E., Meza, C. M., Dumka, L. E., ... & Genalo, M. T. (2007). Engaging Mexican origin families in a school-based preventive intervention. *The Journal of Primary Prevention*, *28*(6), 521–546. doi:

10.1007/s10935-007-0110-z

Grantham-McGregor, S. M., Powell, C. A., Walker, S. P. & Himes, J. H. (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. *The Lancet*, *338*(8758), 1–5. doi: 10.1016/0140-6736(91)90001-6

- Grantham-McGregor, S. M., Walker, S. P., Chang, S. M. & Powell, C. A. (1997). Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children. *The American Journal of Clinical Nutrition*, *66*(2), 247–253.

- Walker, S. P., Chang, S. M., Powell, C. A. & Grantham-McGregor, S. M. (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study. *The Lancet*, *366*(9499), 1804–1807. doi: 10.1016/S0140-6736(05)67574-5

- Walker, S. P., Chang, S. M., Powell, C. A., Simonoff, E. & Grantham-McGregor, S. M. (2006). Effects of psychosocial stimulation and dietary supplementation in early childhood on psychosocial functioning in late adolescence: follow-up of randomised controlled trial. *BMJ*, *333*(7566), 472–476. doi: 10.1136/bmj.38897.555208.2F

- Walker, S. P., Grantham-McGregor, S. M., Powell, C. A. & Chang, S. M. (2000). Effects of growth restriction in early childhood on growth, IQ, and cognition at age 11 to 12 years and the benefits of nutritional supplementation and psychosocial stimulation. *The Journal of Pediatrics*, *137*(1), 36–41. doi: 10.1067/mpd.2000.106227

  - Chang, S. M., Walker, S. P., Grantham-McGregor, S. & Powell, C. A. (2002). Early childhood stunting and later behaviour and school achievement. *Journal of Child Psychology and Psychiatry*, *43*(6), 775–783. doi: 10.1111/1469-7610.00088

Hadeed, J. (2005). *Poverty begins at home: The Mother-Child Education Programme (MOCEP) in the Kingdom of Bahrain.* New York: Peter Lang.

- Hadeed (2005). *The Mother Child Home Education Programme MOCEP - Bahrain.* Power Point presentation Presentation.

- Hadeed, J. (2011). The continued effects of home intervention on child development outcomes in the Kingdom of Bahrain. *Early Child Development and Care*, *181*(10),

1291–1313. doi: 10.1080/03004430.2010.524299

Harris, S. (2010). *Early intervention for poverty-stricken children: A study of preschoolers receiving Jumpstart* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (ProQuest Document ID 1651838299).

High, P. C., LaGasse, L., Becker, S., Ahlgren, I. & Gardner, A. (2000). Literacy promotion in primary care pediatrics: Can we make a difference? *Pediatrics*, *105*(Suppl.3), 927–934.

Jacobs, P. R. (2004). Expository text comprehension instruction for at-risk first-grade children (Doctoral dissertation). *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, *65*(02), 410A.

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B. & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology*, *104*(3), 647–660. doi: 10.1037/a0029018

- Jordan, N. C., Dyson, N. & Glutting, J. (2011). *Developing number sense in kindergartners at risk for learning difficulties in mathematics.* Evanston, IL: Society for Research on Educational Effectiveness. Retrieved from ERIC database (ED528836).

Jungmann, T., Brand, T. & Kurtz, V. (2011). Entwicklungsförderung im Rahmen früher Hilfen. Vorläufige Befunde aus dem Modellprojekt "Pro Kind". *Verhaltenstherapie & Psychosoziale Praxis*, *43*(2), 291–302.

- Adamaszek, K., Brand, T., Kurtz, V. & Jungmann, T. (2013). Stärkung psychischer Ressourcen im Rahmen Früher Hilfen - Erfahrungen und Empfehlungen des Modellprojektes Pro Kind. In K. Makowsky & B. Schücking (Eds.), *Was sagen die Mütter?: Qualitative und quantitative Forschung rund um Schwangerschaft, Geburt und Wochenbett* (pp. 204–224). Weinheim: Beltz, Juventa.

- Brand, T., Kurtz, V. & Jungmann, T. (2012). Qualitätssicherung im Setting Familie – Umsetzungserfahrungen aus dem Modellprojekt "Pro Kind". In *Forschung und Praxis der Gesundheitsförderung, Band 41: Gesund aufwachsen in Kita, Schule, Familie und Quartier. Nutzen und Praxis verhaltens- und verhältnisbezogener Prävention* (pp. 230–237). Köln: BzgA. Retrieved from

http://www.bzga.de/pdf.php?id=3c3429d2892a1cf649de55aa50e3fd6b

Kim, J. S. & Guryan, J. (2010). The efficacy of a voluntary summer book reading intervention for low-income Latino children from language minority families. *Journal of Educational Psychology*, *102*(1), 20–31. doi: 10.1037/a0017270

Kim, J. S., Olson, C. B., Scarcella, R., Kramer, J., Pearson, M., van Dyk, D., ... & Land, R. E. (2011). A randomized experiment of a cognitive strategies approach to text-based analytical writing for mainstreamed Latino English language learners in grades 6 to 12. *Journal of Research on Educational Effectiveness*, *4*(3), 231–263. doi: 10.1080/19345747.2010.523513

Lapointe, C. P. (1977). Effects of motivational and employment programs on scholastic behavior among students from economically disadvantaged groups. *Topics in Culture Learning*, *5*, 50–60.

Larson, K. A. & Rumberger, R. W. (1995). ALAS: Achievement for Latinos through Academic Success. H. Thornton (Ed.), *Staying in school. A technical report of three dropout prevention projects for junior high school students with learning and emotional disabilities.* Minneapolis, MN: University of Minnesota, Institute on Community Integration.

- Rumberger, R. W. & Larson, K. A. (1992). *Keeping high-risk chicano students in school: Lessons from a Los Angeles Middle School Dropout Prevention Program.* Palo Alto, CA: American Institutes for Research in the Behavioral Sciences. Retrieved from ERIC database (ED369582).

Lonigan, C. J., Anthony, J. L., Bloomfield, B. G., Dyer, S. M. & Samwel, C. S. (1999). Effects of two shared-reading interventions on emergent literacy skills of at-risk preschoolers. *Journal of Early Intervention*, *22*(4), 306–322. doi: 10.1177/105381519902200406

Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Brooks-Gunn, J., Boller, K., ... & Berlin, L. J. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. Volume I: Technical report.* Princeton, NJ: Mathematica Policy Research. Retrieved from ERIC database (ED454952).

- Love, J. M. (2010). *Effects of Early Head Start prior to kindergarten entry: The importance*

of early experience. Society for Research on Educational Effectiveness (SREE). Retrieved from ERIC database (ED514662).

- Love, J. M., Chazan-Cohen, R., Raikes, H. & Brooks-Gunn, J. (2013). What makes a difference: Early Head Start evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development*, 78 (1), vii-viii, 1–173. doi: 10.1111/j.1540-5834.2012.00699.x

- Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., ... & Vogel, C. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology*, *41*(6), 885–901. doi: 10.1037/0012-1649.41.6.885

  - Early Head Start Research Consortium. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. Summary report.* Washington, DC. Head Start Bureau: Office of Research and Evaluation., Administration for Children, Youth, and Families (DHHS). Retrieved from ERIC database (ED448894).

  - Fuligni, A. S. & Brooks-Gunn, J. (2013). Mother-child interactions in Early Head Start: Age and ethnic differences in low-income dyads. *Parenting*, *13*(1, SI), 1–26. doi: 10.1080/15295192.2013.732422

  - Habtu, M. A. (2012). *The role of Early Head Start in school readiness* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 1518290).

  - Love, J. M., Constantine, J., Paulsell, D., Boller, K., Ross, C., Raikes, H., ... & Brooks-Gunn, J. (2004). *The role of Early Head Start programs in addressing the child care needs of low-income families with infants and toddlers: Influences on child care use and quality.* Princeton, NJ: Mathematica Policy Research. Retrieved from ERIC database (ED484318).

  - Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Brooks-Gunn, J., Boller, K., ... & Berlin, L. J. (2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families. Volume II: Technical Report Appendixes.* Princeton, NJ: Mathematica Policy Research.

  - Love, J. M., Kisker, E. E., Ross, C. M., Schochet, P. Z., Brooks-Gunn, J., Paulsell, D., ...

& Brady-Smith, C. (2002). *Making a difference in the lives of infants and toddlers and their families: The impacts of Early Head Start. Executive summary.* Princeton, NJ: Mathematica Policy Research.

- Raikes, H. H., Alexander Pan, B., Luze, G., Tamis-LeMonda, C. S., Brooks-Gunn, J., Constantine, J., ... & Rodriguez, E. T. (2006). Mother-child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child Development*, *77*(4), 924–953. doi: 10.1111/j.1467-8624.2006.00911.x

- Raikes, H. H., Brooks-Gunn, J. & Love, J. M. (2013). What makes a difference: Early Head Start evaluation findings in a developmental context: I. Background literature review pertaining to the Early Head Start study. *Monographs of the Society for Research in Child Development*, *78*(1), 1–19. doi: 10.1111/j.1540-5834.2012.00700.x

- Raikes, H. H., Vogel, C. & Love, J. M. (2013). What makes a difference: Early Head Start evaluation findings in a developmental context: IV. Family subgroups and impacts at ages 2, 3, and 5: Variability by race/ethnicity and demographic risk. *Monographs of the Society for Research in Child Development*, *78*(1), 64–92. doi: 10.1111/j.1540-5834.2012.00703.x

- Whiteside-Mansell, L., Bradley, R., McKelvey, L. & Lopez, M. (2009). Center-based Early Head Start and children exposed to family conflict. *Early Education and Development*, *20*(6), 942–957. doi: 10.1080/10409280903206211

Lunkenheimer, E. S., Dishion, T. J., Shaw, D. S., Connell, A. M., Gardner, F., Wilson, M. N. & Skuban, E. M. (2008). Collateral benefits of the Family Check-Up on early childhood school readiness: Indirect effects of parents' positive behavior support. *Developmental Psychology*, *44*(6), 1737–1752. doi: 10.1037/a0013858

- Connell, A., Bullock, B. M., Dishion, T. J., Shaw, D., Wilson, M. & Gardner, F. (2008). Family intervention effects on co-occurring early childhood behavioral and emotional problems: A latent transition analysis approach. *Journal of Abnormal Child Psychology*, *36*(8), 1211–1225. doi: 10.1007/s10802-008-9244-6

- Dishion, T. J., Shaw, D., Connell, A., Gardner, F., Weaver, C. & Wilson, M. (2008). The Family Check-Up with high-risk indigent families: Preventing problem behavior by

increasing parents' positive behavior support in early childhood. *Child Development*, *79*(5), 1395–1414. doi: 10.1111/j.1467-8624.2008.01195.x

- McEachern, A. D., Fosco, G. M., Dishion, T. J., Shaw, D. S., Wilson, M. N. & Gardner, F. (2013). Collateral benefits of the Family Check-Up in early childhood: Primary caregivers' social support and relationship satisfaction. *Journal of Family Psychology*, *27*(2), 271–281. doi: 10.1037/a0031485

  - Dishion, T. J., Nelson, S. E. & Kavanagh, K. (2003). The Family Check-Up with high-risk young adolescents: Preventing early-onset substance use by parent monitoring. *Behavior Therapy*, *34*(4), 553–571. doi: 10.1016/S0005-7894(03)80035-7

  - Dishion, T. J., Shaw, D. S., Connell, A., Gardner, F., Weaver, C. & Wilson, M. (2008). The Family Check-Up with high-risk indigent families: Outcomes of positive parenting and problem behavior from ages 2 through 4 years. *Child Development*, Manuscript under review.

  - Gardner, F., Connell, A., Trentacosta, C. J., Shaw, D. S., Dishion, T. J. & Wilson, M. N. (2009). Moderators of outcome in a brief family-centered intervention for preventing early problem behavior. *Journal of Consulting and Clinical Psychology*, *77*(3), 543–553. doi: 10.1037/a0015622

  - Gardner, F., Hutchings, J., Bywater, T. & Whitaker, C. (2010). Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *Journal of Clinical Child & Adolescent Psychology*, *39*(4), 568–580. doi: 10.1080/15374416.2010.486315

Lynch, S., Kuipers, J., Pyke, C. & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching*, *42*(8), 912–946. doi: 10.1002/tea.20080

- Lynch, S., Taymans, J., Watson, W. A., Ochsendorf, R. J., Pyke, C. & Szesze, M. J. (2007). Effectiveness of a highly rated science curriculum unit for students with disabilities in general education classrooms. *Exceptional Children*, *73*(2), 202–223. doi: 10.1177/001440290707300205

Mac Iver, M. A. (2011). The challenge of improving urban high school graduation outcomes:

Findings from a randomized study of dropout prevention efforts. *Journal of Education for Students Placed at Risk*, *16*(3), 167–184. doi: 10.1080/10824669.2011.584497

Madden, J., O'Hara, J. & Levenstein, P. (1984). Home again: Effects of the Mother-Child Home Program on mother and child. *Child Development*, *55*(2), 636–647. doi: 10.2307/1129975

- Madden, J., Levenstein, P. & Levenstein, S. (1976). Longitudinal IQ outcomes of the mother-child home program. *Child Development*, *47*(4), 1015–1025. doi: 10.2307/1128438

  - Levenstein, P., O'Hara, J. & Madden, J. (1983). The mother-child home program of the verbal interaction project. In Consortium for Longitudinal Studies (Ed.), *As the twig is bent* (pp. 237–263). Hillsdale, NJ: Erlbaum.

Magwaza, A. S. & Edwards, S. D. (1991). An evaluation of an integrated parent-effectiveness training and children's enrichment programme for disadvantaged families. *South African Journal of Psychology*, *21*(1), 21–25. doi: 10.1177/008124639102100103

Massetti, G. M. (2009). Enhancing emergent literacy skills of preschoolers from low-income environments through a classroom-based approach. *School Psychology Review*, *38*(4), 554–569.

Matera, C. & Gerber, M. M. (2008). Effects of a literacy curriculum that supports writing development of Spanish-speaking English learners in Head Start. *NHSA DIALOG*, *11*(1), 25–43. doi: 10.1080/15240750701813865

- Matera, C. (2008). *Effects of writing instruction on Head Start English learners at risk for reading difficulties* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 3335005).

- Matera, C. (2011). Longitudinal outcomes for preschool dual language learners receiving writing instruction. *NHSA Dialog*, *14*(3), 103–120. doi: 10.1080/15240754.2011.586610

Matsumura, L. C., Garnier, H. E. & Spybrook, J. (2013). Literacy coaching to improve student reading achievement: A multi-level mediation model. *Learning and Instruction*, *25*, 35–48. doi: 10.1016/j.learninstruc.2012.11.001

Mayfield, L. G. (2000). *The effects of structured one-on-one tutoring in sight word recognition of first grade students at-risk for reading failure.* Paper presented at the annual meeting of the Mid-South Educational Research Association, Bowling Green, KY. Retrieved from ERIC database (ED449630).

Meier, J. D. & Invernizzi, M. (1999). *Book Buddies in the Bronx: A model for America Reads.* Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. Retrieved from ERIC database (ED436721).

Mendelsohn, A. L., Valdez, P. T., Flynn, V., Foley, G. M., Berkule, S. B., Tomopoulos, S., ... & Dreyer, B. P. (2007). Use of videotaped interactions during pediatric well-child care: Impact at 33 months on parenting and on child development. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *28*(3), 206–212. doi: 10.1097/DBP.0b013e3180324d87

- Mendelsohn, A. L., Dreyer, B. P., Flynn, V., Tomopoulos, S., Rovira, I., Tineo, W., ... & Nixon, A. F. (2005). Use of videotaped interactions during pediatric well-child care to promote child development: A randomized, controlled trial. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *26*(1), 34–41.

Myers, D. E. & Schirm, A. L. (1997). *The national evaluation of Upward Bound. The short-term impact of Upward Bound: An interim report.* Washington, DC: Mathematica Policy Research. Retrieved from ERIC database (ED414849).

- Myers, D. E. & Schirm, A. L. (1999). *The impacts of Upward Bound: Final report for phase I of the national evaluation.* Plainsboro, NJ: Mathematica Policy Research. Retrieved from ERIC database (ED432621).

- Myers, D., Olsen, R., Seftor, N., Young, J. & Tuttle, C. (2004). *The impacts of regular Upward Bound: Results from the third follow-up data collection.* MPR Reference No. 8464-600. Washington, DC: Mathematica Policy Research. Retrieved from ERIC database (ED518667).

  - Maimer, P. J. (2003). *Cultural and social capital transmission in pre-college programs and its impact on the educational attainment of disadvantaged students* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 3092018).

  - Nathan, A. B. (2013). *Does Upward Bound have an effect on student educational outcomes?*

A reanalysis of the Horizons randomized controlled trial study (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 3560135).

Necoechea, D. M. (2007). Children at-risk for poor school readiness: The effect of an early intervention home visiting program on children and parents (Doctoral dissertation). *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, *68*(06), 2311A.

Olds, D. L., Henderson, C. R., Chamberlin, R. & Tatelbaum, R. (1986). Preventing child abuse and neglect: A randomized trial of nurse home visitation. *Pediatrics*, *78*(1), 65–78.

- Olds, D. L., Henderson Jr, C. R., Phelps, C., Kitzman, H. & Hanks, C. (1993). Effect of prenatal and infancy nurse home visitation on government spending. *Medical Care*, *31*(2), 155–174.

- Olds, D. L., Henderson Jr, C. R., Tatelbaum, R. & Chamberlin, R. (1988). Improving the life-course development of socially disadvantaged mothers: A randomized trial of nurse home visitation. *American Journal of Public Health*, *78*(11), 1436–1445. doi: 10.2105/AJPH.78.11.1436

- Olds, D. L., Henderson, C. R. & Kitzman, H. (1994). Does prenatal and infancy nurse home visitation have enduring effects on qualities of parental caregiving and child health at 25 to 50 months of life?. *Pediatrics*, *93*(1), 89–98.

  - Olds, D. L., Eckenrode, J., Henderson, C. R., Kitzman, H., Powers, J., Cole, R., ... & Luckey, D. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: Fifteen-year follow-up of a randomized trial. *JAMA*, *278*(8), 637–643. doi: 10.1001/jama.1997.03550080047038

  - Olds, D., Henderson Jr, C. R., Cole, R., Eckenrode, J., Kitzman, H., Luckey, D., ... & Powers, J. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *JAMA*, *280*(14), 1238–1244. doi: 10.1001/jama.280.14.1238

  - Olds, D., Henderson, C. R., Kitzman, H. & Cole, R. (1995). Effects of prenatal and infancy nurse home visitation on surveillance of child maltreatment. *Pediatrics*, *95*(3), 365–372.

- Olds, D., Henderson, C. R., Kitzman, H., Eckenrode, J., Cole, R., Tatelbaum, D. L. O. R., ..., & Hill, R. C. P. (1998). *Prenatal and infancy home visitation by nurses: A program of research.*College Park, MD: Conference Proceeding of the Preventing Second Births to Teenage Mothers conference.

- Zielinski, D. S., Eckenrode, J. & Olds, D. L. (2009). Nurse home visitation and the prevention of child maltreatment: Impact on the timing of official reports. *Development and Psychopathology*, *21*(2), 441–453. doi: 10.1017/S0954579409000248

Palmer, J. C. (2007). The efficacy of planetarium experiences to teach specific science concepts (Doctoral dissertation). *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, *68*(03), 939A.

Pantin, H., Coatsworth, J. D., Feaster, D. J., Newman, F. L., Briones, E., Prado, G., ... & Szapocznik, J. (2003). Familias Unidas: The efficacy of an intervention to promote parental investment in Hispanic immigrant families. *Prevention Science*, *4*(3), 189–201. doi: 10.1023/A:1024601906942

Penuel, W. R., Bates, L., Gallagher, L. P., Pasnik, S., Llorente, C., Townsend, E., ... & VanderBorght, M. (2012). Supplementing literacy instruction with a media-rich intervention: Results of a randomized controlled trial. *Early Childhood Research Quarterly*, *27*(1), 115–127. doi: 10.1016/j.ecresq.2011.07.002

Peterson, C. Q. (1999). *The effect of an occupational therapy intervention on handwriting in academically at-risk first graders* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9946543)

- Peterson, C. Q., & Nelson, D. L. (2003). Effect of an occupational intervention on printing in children with economic disadvantages. *American Journal of Occupational Therapy*, *57*(2), 152–160. doi: 10.5014/ajot.57.2.152

Plasencia-Peinado, J. (1998). Examining the effect of three Spanish reading approaches on the reading achievement of first-grade limited English proficient (LEP) students at-risk for special education (Doctoral dissertation). *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, *59*(07), 2311A.

Powell, D. R., Diamond, K. E., Burchinal, M. R. & Koehler, M. J. (2010). Effects of an early

literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology*, *102*(2), 299–312. doi: 10.1037/a0017763

Powell, C. & Grantham-McGregor, S. (1989). Home visiting of varying frequency and child development. *Pediatrics*, *84*(1), 157–164.

Powers, S. & Price-Johnson, C. (2007). *Evaluation of the Waterford Early Math & Science Program for Kindergarten: First-Year Implementation in Five Urban Low-Income Schools.* Online Submission, retrieved from ERIC database (ED501575).

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ... & Spier, E. (2010). *Head Start impact study. Final report.* Administration for Children and Families (ACF), Office of Planning, Research & Evaluation, Westat, Inc. Revtrieved from ACF website: http://www.acf.hhs.gov/sites/default/files/opre/executive_summary_final.pdf

- Lipscomb, S. T., Pratt, M. E., Schmitt, S. A., Pears, K. C. & Kim, H. K. (2013). School readiness in children living in non-parental care: Impacts of Head Start. *Journal of Applied Developmental Psychology*, *34*(1), 28–37. doi: 10.1016/j.appdev.2012.09.001

- Puma, M., Bell, S., Cook, R., Heid, C., Lopez, M., Zill, N., … & Bernstein, H. (2005). *Head Start impact study: First year findings.* Administration for Children and Families (ACF), Office of Planning, Research & Evaluation, Westat, Inc. Revtrieved from ACF website: http://www.acf.hhs.gov/sites/default/files/opre/first_yr_finds.pdf

- Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ... & Spier, E. (2010). *Head Start impact study. Final report. Executive summary.* Administration for Children and Families (ACF), Office of Planning, Research & Evaluation, Westat, Inc. Revtrieved from ACF website: http://www.acf.hhs.gov/sites/default/files/opre/executive_summary_final.pdf

Ramani, G. B. & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*(2), 375–394. doi: 10.1111/j.1467-8624.2007.01131.x

Ramey, C. T., Bryant, D. M., Campbell, F. A., Sparling, J. J. & Wasik, B. H. (1988). Early intervention for high-risk children: The Carolina early intervention program. In R. H. Price, E. L. Cowen, R. P. Lorion, & J. Ramos-McKay (Eds.), *Fourteen ounces of prevention:*

A casebook for practitioners (pp. 32–43). Washington, DC, US: American Psychological Association. doi: 10.1037/10064-003

- Campbell, F. A. & Pungello, E. (2000). *High quality child care has long-term educational benefits for poor children.* Paper presented at the 5th Head Start National Research Conference, Washington, DC. Retrieved from ERIC database (ED443545).

- Campbell, F. A. & Ramey, C. T. (1993). *Mid-adolescent outcomes for high risk students: An examination of the continuing effects of early intervention.* Paper presented at the 60th Biennial Meeting of theSociety for Research in Child Development, New Orleans, LA. Retrieved from ERIC database (ED358919).

- Campbell, F. A. & Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: a follow-up study of children from low-income families. *Child Development*, *65*(2), 684–698. doi: 10.1111/j.1467-8624.1994.tb00777.x

- Campbell, F. A. & Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal*, *32*(4), 743–772. doi: 10.3102/00028312032004743

- Campbell, F. A., Helms, R., Sparling, J. J. & Ramey, C. T. (1998). Early-childhood programs and success in school: The Abecedarian study. In W. S. Barnett & S. S. Boocock (Eds.), *Early care and education for children in poverty: Promises, programs, and long-term results* (pp. 145–166). Albany, NY, US: State University of New York Press.

- Campbell, F. A., Pungello, E. P., Burchinal, M., Kainz, K., Pan, Y., Wasik, B. H., ... & Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up. *Developmental Psychology*, *48*(4), 1033–1043. doi: 10.1037/a0026644

- Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M. & Ramey, C. T. (2001). The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Developmental Psychology*, *37*(2), 231–242. doi: 10.1037/0012-1649.37.2.231

- Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J. & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied*

Developmental Science, *6*(1), 42–57. doi: 10.1207/S1532480XADS0601_05

- Campbell, F. A., Wasik, B. H., Pungello, E., Burchinal, M., Barbarin, O., Kainz, K., ... & Ramey, C. T. (2008). Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly*, *23*(4), 452–466. doi: 10.1016/j.ecresq.2008.03.003

- McLaughlin, A. E., Campbell, F. A., Pungello, E. P. & Skinner, M. (2007). Depressive symptoms in young adults: The influences of the early home environment and early educational child care. *Child Development*, *78*(3), 746–756. doi: 10.1111/j.1467-8624.2007.01030.x

- Ramey, C. T. & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency*, *88*(5), 515–523.

- Ramey, C. T., Collier, A. M., Sparling, J. J., Loda, F. A., Campbell, F. A., Ingram, D. L. & Finkelstein, N. W. (1974): T*he Carolina Abecedarian Project: A longitudinal and multidisciplinary approach to the prevention of developmental retardation.* Chapel Hill, CA: Frank Porter Graham Child Development Center, University of North Carolina. Retrieved from ERIC database (ED104548).

- Ramey, C.T. & Campbell, F.A. (1991). Poverty, early childhood education, and academic competence: The Abecedarian experiment. In A.C. Huston (Ed.), *Children in poverty: Child development and public policy* (pp. 190–221). Cambridge, England: Cambridge University Press.

  - Boat, B. W., Campbell, F. A. & Ramey, C. T. (1986). Preventive education and birth order as co-determinants of IQ in disadvantaged 5-year-olds. *Child: Care, Health & Development*, *12*(1), 25–36. doi: 10.1111/j.1365-2214.1986.tb00484.x

  - Campbell, F. A. & Ramey, C. T. (2007, December). *Carolina Abecedarian Project.* Paper presented at the National Invitational Conference of the Early Childhood Research Collaborative, Minneapolis, MN.

  - Campbell, F. A., Goldstein, S., Schaefer, E. S. & Ramey, C. T. (1991). Parental beliefs and values related to family risk, educational intervention, and child academic competence.

Early Childhood Research Quarterly, *6*(2), 167–182. doi: 10.1016/0885-2006(91)90005-6

- Campbell, F. & Ramey, C. (1977). *The effects of early intervention on intellectual development.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, LA.

- Feagans, L. V. & Farran, D. C. (1994). The effects of daycare intervention in the preschool years on the narrative skills of poverty children in kindergarten. *International Journal of Behavioral Development*, *17*(3), 503–523. doi: 10.1177/016502549401700307

- Feagans, L. V., Fendt, K. & Farrn, D. C. (1995). The effects of day care intervention on teacher's ratings of the elementary school discourse skills in disadvantaged children. *International Journal of Behavioral Development*, *18* (2), 243–261. doi: 10.1177/016502549501800204

- Martin, S. L., Ramey, G. T. & Ramey, S. L. (1990). The prevention of intellectual impairment in children of impoverished families: Findings of a randomized trial of educational daycare. *American Journal of Public Health*, *80*(7), 844–847.

- Masse, L. N. & Barnett, S. W. (2002). *A benefit-cost analysis of the Abecedarian Early Childhood Intervention*. New Brunswick, NJ: National Institute for Early Education Research (NIEER). Retrieved from the NIEER website: http://nieer.org/resources/research/AbecedarianStudy.pdf

- Ramey, C. T. (1992). High-risk children and IQ: altering intergenerational patterns. *Intelligence*, *16*(2), 239–256. doi: 10.1016/0160-2896(92)90007-E

- Ramey, C. T. & Campbell, F. A. (1981). Educational intervention for children at risk for mild retardation: A longitudinal analysis. In P. Mittler (Ed.), *Frontiers of knowledge in mental retardation: Vol. 1. Social, education, and behavioral aspects* (pp. 47–57). Baltimore, MD: University Park Press.

- Ramey, C. T. & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency*, *88*(5), 515–523.

- Ramey, C. T. & Campbell, F. A. (1987). The Carolina Abecedarian Project: An educa-

tional experiment concerning human malleability. In J. Gallagher & C. Ramey (Eds.), *The malleability of children* (pp. 115–126). Baltimore: Brookes.

◦ Ramey, C. T. & Farran, D. C. (1983). *Intervening with high-risk families via infant day-care*. Paper presented at the 50th Biennial Meeting of the Society for Research in Child Development, Detroit, MI. Retrieved from ERIC database (ED230289).

◦ Ramey, C. T. & Gowen, J. W. (1984). A general systems approach to modifying risk for retarded development. *Early Child Development and Care*, *16*(1-2), 9–26. doi: 10.1080/0300443840160102

◦ Ramey, C. T. & Haskins, R. (1981). The causes and treatment of school failure: Insights from the Carolina Abecedarian Project. In M. Begab, H. Garber, & H. G. Haywood (Eds.), *Causes and prevention of retarded development in psychosocially disadvantaged children* (pp. 89–112). Baltimore: University Park Press.

◦ Ramey, C. T. & Ramey, S. L. (1998). Prevention of intellectual disabilities: Early interventions to improve cognitive development. *Preventive Medicine*, *27*(2), 224–232. doi: 10.1006/pmed.1998.0279

◦ Ramey, C. T., Bryant, D. M., Campbell, F. A., Sparling, J. J. & Wasik, B. H. (1989). Early intervention for high-risk children: The Carolina early intervention program. *Prevention in Human Services*, *7*(1), 33–57. doi: 10.1300/J293v07n01_03

◦ Ramey, C. T., Campbell, F. A., Burchinal, M., Skinner, M. L., Gardner, D. M. & Ramey, S. L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science*, *4*(1), 2–14. doi: 10.1207/S1532480XADS0401_1

◦ Ramey, C. T., Dorval, B. & Baker-Ward, L. (1983). Group day care and socially disadvantaged families: Effects on the child and the family. *Advances in Early Education & Day Care*, *3*, 69–106.

◦ Ramey, C. T., Farran, D. C. & Campbell, F. A. (1979). Predicting IQ from mother-infant interactions. *Child Development*, *50*(3), 804–814. doi: 10.2307/1128947

◦ Ramey, C. T., MacPhee, D. & Yeates, K. O. (1982). Preventing developmental retardation: A general systems model. In J. M. Joffee & L. A. Bond (Eds.), *Facilitating infant*

and early childhood development (pp. 343–401). Hanover, NH: University Press of New England.

- Ramey, C. T., Yeates, K. O. & MacPhee, D. (1984). Risk for retarded development among disadvantaged families: A systems theory approach to preventive intervention. In A. F. Rotatori, B. K. Keogh, E. C., F. W. Litton, & J. Gottlieb (Eds.), *Advances in Special Education: Documenting program impact* (Vol. 4, pp. 249–272). Greenwich, CT: JAI Press.

- Ramey, C. T., Yeates, K. O. & Short, E. J. (1984). The plasticity of intellectual development: Insights from preventive intervention. *Child Development*, *55*(5), 1913–1925. doi: 10.2307/1129938

- Ramey, G. T., McGinness, G. D., Gross, L., Gollier, A. M. & Barrie-Blackley, S. (1982). The Abecedarian approach to social competence: Cognitive and linguistic intervention for disadvantaged preschoolers. In K. Borman (Ed.), *The social life of children in a changing society* (pp. 145–174). Hillsdale, NJ: Erlbaum Associates.

Ramey, C. T. & Haskins, R. (1981). The modification of intelligence through early experience. *Intelligence*, *5*(1), 5–19. doi: 10.1016/0160-2896(81)90013-1

Ransford-Kaldon, C., Flynt, E. S. & Ross, C. (2011). *A randomized controlled trial of a response-to-intervention (RTI) Tier 2 literacy program: Leveled Literacy Intervention (LLI).* Society for Research on Educational Effectiveness (SREE). Retrieved from ERIC database (ED518772).

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M. & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*, *48*(2), 462–501. doi: 10.3102/0002831210382891

Rhodes, R. L., Whitten, J. D. & Copeland, E. P. (1997). Early intervention with at-risk Hispanic students: Effectiveness of the Piacceleration program in developing Piagetian intellectual processes. *Journal of Experimental Education*, *65*(4), 318–328. doi: 10.1080/00220973.1997.10806607

Riksen-Walraven, J. M. (1978). Effects of caregiver behavior on habituation rate and self-efficacy in infants. *International Journal of Behavioral Development*, *1*(2), 105–130. doi: 10.1177/016502547800100202

Ritter, G. W. (2000). *The academic impact of volunteer tutoring in urban public elementary schools: Results of an experimental design evaluation* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 9965556).

Roggman, L. A., Boyce, L. K. & Cook, G. A. (2009). Keeping kids on track: Impacts of a parenting-focused Early Head Start program on attachment security and cognitive development. *Early Education and Development*, *20*(6), 920–941. doi: 10.1080/10409280903118416

Rutherford, T., Kibrick, M., Burchinal, M., Richland, L., Conley, A., Osborne, K., ... & Martinez, M. E. (2010). *Spatial temporal mathematics at scale: An innovative and fully developed paradigm to boost math achievement among all learners.* Paper presented at the 91st annual meeting of the American Educational Research Association, Denver, CO. Online Submission. Retrieved from ERIC database (ED510612).

- Kibrick, M., Rutherford, T., Burchinal, M. R., Richland, L. E., Conley, A., Long, J. J., ... & Martinez, M. (2010). *The effects of ST Math on standardized test scores: A randomized field study.* Retrieved from http://www.education.uci.edu/CRCL/documents/IES\%202010\%20Poster.pdf

Scarr, S. & McCartney, K. (1988). Far from home: An experimental evaluation of the Mother-Child Home Program in Bermuda. *Child Development*, *59*(3), 531–543. doi: 10.2307/1130555

Schacter, J. & Jo, B. (2005). Learning when school is not in session: A reading summer day-camp intervention to improve the achievement of exiting first-grade students who are economically disadvantaged. *Journal of Research in Reading*, *28*(2), 158–169. doi: 10.1111/j.1467-9817.2005.00260.x

Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., ... & Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, *104*(4), 591–618. doi: 10.1037/a0031495

Shmukler, D. & Naveh, I. (1985). Structured vs. unstructured play training with economically disadvantaged preschoolers. *Imagination, Cognition and Personality*, *4*(3), 293–304. doi:

10.2190/WQ65-UNAM-VDV9-PLMR

Slaughter, D. T. (1979). *Longitudinal assessment of the intelligence of black infants, ages 22 to 41 months.* Revised (NIH Publication No. BBB-00456). National Institute of Child and Human Development Bethesda, MD. Retrieved from ERIC databse (ED199358).

- Slaughter, D. T., Earls, F. & Caldwell, B. M. (1983). Early intervention and its effects on maternal and child development. *Monographs of the Society for Research in Child Development*, *48*(4), 1–83. doi: 10.2307/1165896

St. Pierre, R. G. S., Layzer, J. I., Goodson, B. D. & Berstein, L. S. (1997). *National impact evaluation of the Comprehensive Child Development Program: Final report.* Cambridge, MA: Abt Associates Inc.. Retrieved from the Abt Associates website: www.abtassociates.com/reports/D19970050.pdf

- Goodson, B. D., Layzer, J. I., St. Pierre, R. G., Bernstein, L. S. & Lopez, M. (2000). Effectiveness of a comprehensive, five-year family support program for low-income children and their families: Findings from the Comprehensive Child Development Program. *Early Childhood Research Quarterly*, *15*(1), 5–39. doi: 10.1016/S0885-2006(99)00040-X

- Ryan, C. S., McCall, R. B., Robinson, D. R., Groark, C. J., Mulvey, L. & Plemons, B. W. (2002). Benefits of the Comprehensive Child Development Program as a function of AFDC receipt and SES. *Child Development*, *73*(1), 315–328. doi: 10.1111/1467-8624.00408

- St. Pierre, R. G. S. & Layzer, J. I. (1999). Using home visits for multiple purposes: The Comprehensive Child Development Program. *Future of Children*, *9*(1), 134–151. doi: 10.2307/1602725

- St. Pierre, R. G. S., Layzer, J. I., Goodson, B. D. & Bernstein, L. S. (1997). *National impact evaluation of the Comprehensive Child Development Program: Executive summary.* Cambridge, MA: Abt Associates Inc.. Retrieved from the Administration for Children & Families website: http://acf.hhs.gov/sites/default/files/opre/ccdp00_natl.pdf

- St. Pierre, R. G. S., Layzer, J. I., Goodson, B. D. & Bernstein, L. S. (1997). *The effectiveness of comprehensive case management interventions: Findings from the national evaluation of the Comprehensive Child Development Program.* Cambridge, MA: Abt Associates Inc.. Retrieved from the Abt Associates website: www.abtassociates.com/reports/paper6.pdf

- St. Pierre, R. G. S., Layzer, J. I., Goodson, B. D. & Bernstein, L. S. (1999). The effectiveness of comprehensive, case management interventions: Evidence from the national evaluation of the Comprehensive Child Development Program. *The American Journal of Evaluation*, *20*(1), 15–34. doi: 10.1177/109821409902000103

St. Pierre, R. G. S., Ricciuti, A. E., Tao, F., Creps, C., Swartz, J. P., Lee, W., ... & Rimdzius, T. A. (2003). *Third national Even Start evaluation: Program impacts and implications for improvement.* Cambridge, MA & Bethesda, MD: Abt Associates Inc.. Retrieved from ERIC database (ED475962).

- Ricciuti, A. E., Pierre, R. S., Lee, W., Parsad, A. & Rimdzius, T. A. (2004). *Third national Even Start evaluation: Follow-up findings from the experimental design study.* Cambridge, MA & Bethesda, MD: Abt Associates Inc.. Retrieved from ERIC databse (ED475962).

- St. Pierre, R. G. S., Ricciuti, A. E. & Rimdzius, T. A. (2005). Effects of a family literacy program on low-literate children and their parents: Findings from an evaluation of the Even Start family literacy program. *Developmental Psychology*, *41*(6), 953–970. doi: 10.1037/0012-1649.41.6.953

St. Pierre, R. G. S., Swartz, J. P., Gamse, B. C., Murray, S., Deck, D. & Nickel, P. (1995). *National evaluation of the Even Start Family Literacy Program: Final report.* RMC Research Corp., Arlington, VA, Abt Associates Inc., Bethesda, MD. Retrieved from ERIC database (ED386328).

- Gamse, B. C., Conger, D., Elson, D. & McCarthy, M. (1997). *Follow-up study of families in the even start in-depth study. Final report.* Cambridge, MA: Abt Associates Inc.. Retrieved from the Abt Associates website: www.abtassociates.com/reports/D19970044.pdf

- Tao, F., Gamse, B. C. & Tarr, H. (1998). *National Evaluation of the Even Start Family Literacy Program: 1994-1997 Final Report.* Bethesda, MD: Fu Associates, Ltd., Abt Associates Inc.. Retrieved from ERIC database (ED427889).

Tan, J. P., Lane, J. & Lassibille, G. (1999). Student outcomes in Philippine elementary schools: An evaluation of four experiments. *The World Bank Economic Review*, *13*(3), 493–508. doi: 10.1093/wber/13.3.493

Taylor, A. L. (2010). *Testing a model of change in achievement mentoring for school behavior problems* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 3434857).

- Clarke, L. O. (2009). Effects of a school-based adult mentoring intervention on low income, urban high school freshmen judged to be at risk for drop-out: A replication and extension (Doctoral dissertation). *Dissertation Abstracts International: Section B. The Sciences and Engineering*, *70*(11), 7232B.

- Holt, L. J., Bry, B. H. & Johnson, V. L. (2008). Enhancing school engagement in at-risk, urban minority adolescents through a school-based, adult mentoring intervention. *Child & Family Behavior Therapy*, *30*(4), 297–318. doi: 10.1080/07317100802482969

Taylor, D. K. (1991). *Primary prevention of infant growth and development dysfunction: The promotion of positive parenting practices* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database (UMI No. 9208855).

Tazi, Z. (2012). The effects of bilingual instruction on the English emergent literacy skills of Spanish-speaking preschool children (Doctoral dissertation). *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, *72*(08), 2666A.

Tolan, P., Gorman-Smith, D., & Henry, D. (2004). Supporting families in a high-risk setting: Proximal effects of the SAFEChildren preventive intervention. *Journal of Consulting & Clinical Psychology*, *72*(5), 855–869. doi: 10.1037/0022-006X.72.5.855

Tucker, C. M., Chennault, S. A., Brady, B. A., Fraser, K. P., Gaskin-Butler, V. T., Dunn, C. & Frisby, C. (1995). A parent, community, public schools, and university involved partnership education program to examine and boost academic achievement and adaptive functioning skills of African American students. *Journal of Research and Development in Education*, *28*(3), 174–185.

- Tucker, C. M., Herman, K. C., Reid, A. D., Keefer, N. L. & Vogel, D. L. (1998). The research-based Model Partnership Education Program: A 4-year outcome study. *Journal of Research & Development in Education*, *32*(1), 32–37.

Tuttle, C. C., Gill, B., Gleason, P., Knechtel, V., Nichols-Barrer, I. & Resch, A. (2013). *KIPP*

middle schools: Impacts on achievement and other outcomes. Final report. Mathematica Policy Research, Inc. Retrieved from ERIC database (ED540912).

Uchikoshi, Y. (2005). Narrative development in bilingual kindergarteners: Can Arthur help? *Developmental Psychology*, *41*(3), 464–478. doi: 10.1037/0012-1649.41.3.464

- What Works Clearinghouse (2006). *Arthur. What Works Clearinghouse Intervention Report.* Retrieved from ERIC database (ED510612).

Vadasy, P. F. & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal of Educational Psychology*, *102*(4), 786–803. doi: 10.1037/a0019639

- Vadasy, P. F. & Sanders, E. A. (2012). Two-year follow-up of a kindergarten phonics intervention for English learners and native English speakers: Contextualizing treatment impacts by classroom literacy instruction. *Journal of Educational Psychology*, *104*(4), 987–1005. doi: 10.1037/a0028163

Vadasy, P. F. & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading*, *15*(6), 471–497. doi: 10.1080/10888438.2010.501091

- Vadasy, P. F. & Sanders, E. A. (2013). Two-year follow-up of a code-oriented intervention for lower-skilled first-graders: Zhe influence of language status and word reading skills on third-grade literacy outcomes. *Reading and Writing*, *26*(6), 821–843. doi: 10.1007/s11145-012-9393-4

Wagner, M., Spiker, D., Hernandez, F., Song, J. & Gerlach-Downie, S. (2001). *Multisite Parents as Teachers evaluation: Experiences and outcomes for children and families.* Menlo Park, CA: SRI International.

  - Wagner, M., Spiker, D. & Linn, M. I. (2002). The effectiveness of the Parents as Teachers program with low-income parents and children. *Topics in Early Childhood Special Education*, *22*(2), 67–81. doi: 10.1177/02711214020220020101

Wang, H. & Woodworth, K. (2011). *A randomized controlled trial of two online mathematics curricula.* Society for Research on Educational Effectiveness (SREE). Retrieved from ERIC database (ED528686).

Weikart, D. P., Deloria, D. J., Lawser, S. A. & Wiegerink, R. (1970). *Longitudinal results of the Ypsilanti Perry Preschool Project. Final report. Volume II of 2 volumes.* Ypsilanti, MI: High/Scope Educational Research Foundation. Retrieved from ERIC database (ED044536).

- Berrueta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S. & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19.* Ypsilanti, MI: High/Scope Educational Research Foundation. Retrieved from ERIC database (ED313128).

- Muennig, P., Schweinhart, L., Montie, J. & Neidell, M. (2009). Effects of a prekindergarten educational intervention on adult health: 37-year follow-up results of a randomized controlled trial. *American Journal of Public Health*, *99*(8), 1431–1437. doi: 10.2105/AJPH.2008.148353

- Schweinhart, L. J. (2013). Long-term follow-up of a preschool experiment. *Journal of Experimental Criminology*, *9*(4), 389–409. doi: 10.1007/s11292-013-9190-3

- Schweinhart, L. J. & Weikart, D. P. (1980). *Young children grow up: The effects of the Perry Preschool Program on youths through age 15.* Ypsilanti, MI: High/Scope Educational Research Foundation.

- Schweinhart, L. J. & Weikart, D. P. (1981). Effects of the Perry Preschool Program on youths through age 15. *Journal of Early Intervention*, *4*(1), 29–39. doi: 10.1177/105381518100400105

- Schweinhart, L. J., Berrueta-Clement, J. R., Barnett, W. S., Epstein, A. S. & Weikart, D. P. (1985). Effects of the Perry Preschool program on youths through age 19: A summary. *Topics in Early Childhood Special Education*, *5*(2), 26–35. doi: 10.1177/027112148500500204

- Weikart, D. P., Rogers, L., Adcock, C. & McClelland, D. (1970). *The cognitively oriented curriculum: A framework for preschool teachers. Final report. Volume I of 2 volumes.* Ypsilanti, MI: High/Scope Educational Research Foundation. Retrieved from ERIC

database (ED044535).

- Nores, M., Belfield, C. R., Barnett, W. S. & Schweinhart, L. (2005). Updating the economic impacts of the High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis*, *27*(3), 245–261. doi: 10.3102/01623737027003245

- Schweinhart, L. J. (2003). *Benefits, costs, and explanation of the High/Scope Perry Preschool Program.* Paper presented at the 2003 meeting of the Society for Research in Child Development, Tampa, FL. Retrieved from ERIC database (ED475597).

- Schweinhart, L. J. (2010). The challenge of the HighScope Perry Preschool study. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, & J. A. Temple (Eds.), *Childhood programs and practices in the first decade of life: A human capital integration* (pp. 157–167). New York: Cambridge University Press.

- Schweinhart, L. J. & Weikart, D. P. (1993). Success by empowerment: The High/Scope Perry Preschool study through age 27. *Young Children*, *49*(1), 54–58.

- Schweinhart, L. J. & Weikart, D. P. (1997). The High/Scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly*, *12*(2), 117–143. doi: 10.1016/S0885-2006(97)90009-0

- Schweinhart, L. J., Barnes, H. V. & Weikart, D. P. (1993). Significant benefits: The High/Scope Perry Preschool study through age 27 (Summary of Monograph). *American Youth Policy Forum*, 134–136.

- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R. & Nores, M. (2005). *The High/Scope Perry Preschool study through age 40: Summary, conclusions, and frequently asked questions (Summary of the monograph)*. Ypsilanti, MI: High/Scope Press.

- Schweinhart, L. J., Weikart, D. P. & Larner, M. B. (1986). Child-initiated activities in early childhood programs may help prevent delinquency. *Early Childhood Research Quarterly*, *1*(3), 303–312. doi: 10.1016/0885-2006(86)90038-4

- Schweinhart, L. L., Weikart, D. P., & Larner, M. B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly*, *1*(1), 15–46. doi: 10.1016/0885-2006(86)90005-0

- Schweinhart, L.J. & Xiang, Z. (2003). *Evidence that the High/Scope Perry Preschool Program prevents adult crime.* Paper presented at the 2003 American Society of Criminology Conference, Denver, CO.

- Schweinhart, L.J., (2007). Crime Prevention by the High/Scope Perry Preschool Program. *Victims and Offenders*, *2*(2), 141–160. doi: 10.1080/15564880701263064

- Weikart, D. P. & Schweinhart, L. J. (1992). High/Scope Preschool program outcomes. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 67–86). New York, NY: The Guilford Press.

White, T. G. & Kim, J. S. (2008). Teacher and parent scaffolding of voluntary summer reading. *The Reading Teacher*, *62*(2), 116–125. doi: 10.1598/RT.62.2.3

- Kim, J. S. (2006). Effects of a voluntary summer reading intervention on reading achievement: Results from a randomized field trial. *Educational Evaluation and Policy Analysis*, *28*(4), 335–355. doi: 10.3102/01623737028004335

- Kim, J. S. & White, T. G. (2008). Scaffolding voluntary summer reading for children in grades 3 to 5: An experimental study. *Scientific Studies of Reading*, *12*(1), 1–23. doi: 10.1080/10888430701746849

Widmayer, S., Peterson, L., Calderon, A., Carnaham, S. & Wingerd, J. L. (1992). The Haitian Perinatal Intervention Project. In M. Larner, R. Halpern & O. Harkavy (Eds.), *Fair Start for Children: Lessons learned from seven demonstration projects* (pp. 115–135). New Haven, CT: Yale University Press.

- Widmayer, S. M. (1988). *Intervention with Haitian entrants in South Florida.* Fort Lauderdale, FL: Children's Diagnostic and Treatment Center. Retrieved from ERIC database (ED314163).

Wilkins, C., Gersten, R., Decker, L. E., Grunden, L., Brasiel, S., Brunnert, K. & Jayanthi, M. (2012). *Does a summer reading program based on Lexiles affect reading comprehension? Final report. NCEE 2012-4006.* National Center for Education Evaluation and Regional Assistance. Retrieved from ERIC database (ED530029).

Wynne, C. B. (2011). The implementation of a promising self-affirmation intervention among

negatively stereotyped students in a randomized field experiment (Doctoral dissertation). *Dissertation Abstracts International: Section A. Humanities and Social Sciences*, *73*(02), 475A.

H Coding Handbook

The effectiveness of psychosocial and educational programs for preventing or compensating the negative consequences in the development of children and adolescents caused by poverty or migration: A multinational meta-analysis.

Coding Handbook

Technical information

- **Variable type ☐/☒:** This type is a dummy variable. Check the ones that apply. Multiple selections within variable are allowed.
- **Hierarchy:** The number written in brackets behind a variable serves as element to structure the next variable, e.g. if program orientation's dummy variable child/adolescent promotion (1) is checked, then at least one of the program aim variables marked with a "1" (prior to the variable) needs to be checked, too. Also applies for the addressee and program content/focus.

Abbreviations

ES = effect size, *M* = mean, *SD* = standard deviation, *N* = sample size, *IG* = intervention group, *CG* = control group

DOCUMENT LEVEL

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Publication ID	First four letters of the first authors' last name + first letter of the second and third authors last names + last two digits of the publication year + A/M/B for the study type (poverty/migration/both), e.g. poverty study from **McCl**anahan, Wendy S., **S**ipe, Cynthia L., **S**mith, Thomas J. (20**04**) → McClSS04A Note: Omit letters when there are fewer than four authors, e.g. Ramey, Craig T. & Haskins, Ron (1981) → RameH81A	ID_Pub
Date coded	Present date (DD.MM.JJJJ)	date
Coder	Name of coder	coder
Author(s)	Write out the first six authors, then use "et al."; omit (initials of) surnames	authors
Title	Main title only	title
Publication year	JJJJ	pub_year
Publication format	1 = **Peer-reviewed journal article** (published; article published in a peer reviewed journal) 2 = **Journal article with unknown peer-review status** (published, article published in a journal whose peer review status is unknown) 3 = **Published scientific report** (published; scientific report that didn't undergo comprehensive independent peer review before publication; e.g. publication of a ministry) 4 = **Unpublished scientific report** (unpublished; any scientific work that is not publicly available (e.g. via internet or the ThULB) and only obtainable by contacting the author) 5 = **Doctoral thesis** (unpublished; Thesis for obtaining a doctor's degree) 6 = **Conference proceedings** (published; collection of academic papers published in the context of an academic conference) 7 = **Book** (published; scientific report published in form of a book) 8 = **Book chapter** (published; any scientific report published in a book) See appendix A for examples of peer-/not peer-reviewed journals	pub_form
Name of journal	Fill out only, if "Publication format" = 1 or 2, otherwise leave blank	journal
Publication language	1 = English 2 = German 3 = Other	lang
Publication language: other	Fill out only if "Publication language" = 3, otherwise leave blank	lang_o

Country of performance	1 = USA 2 = UK 3 = Canada 4 = Netherlands 5 = Germany 6 = Belgium 7 = Sweden 8 = Norway 9 = Other Country where the study took place; if no other information available, use first authors' address	countr
Country of performance: other	Fill out only if "Country of performance" = 9, otherwise leave blank	countr_o
Additional literature	0 = **Only information from original study was used for coding the publication** (no additional literature was used) 1 = **Supplemental duplicate(s)** was/were used for coding (i.e. an eligible study referring to the same analysis and content as the original study, but contains additional information, e.g. further outcomes) 2 = **External publication(s)** was/were used for coding (i.e. a publication that did not meet the inclusion criteria, like a publication of the program curriculum or the sample description) 3 = **Both**: supplemental and external publications were used for coding. See appendix B for more information concerning literature types	add_lit
Name supplemental duplicate(s)	Fill out only if "Additional literature" = 1 or 3, otherwise leave blank Note complete citation.	sup_dub_tx
Name external publication(s)	Fill out only if "Additional literature" = 2 or 3, otherwise leave blank Note complete citation.	ext_pub_tx

COMPARISON LEVEL

INFORMATION ABOUT THE CHARACTERISTICS OF THE PROGRAM

VARIABLE		CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Publication ID			ID_Pub
Comparison ID		Number of comparison; if only subgroups are being analysed, treat them as different comparisons	ID_Comp
BASIC PROGRAM INFORMATION			
Program name			pr_name
Program orientation *Who is supposed to benefit from the program according to the author?*		☐/☒ **Child/Adolescent promotion** (1)	pr_or_ch
		☐/☒ **Parent Training/Family promotion** (2)	pr_or_pf
		☐/☒ **Environmental change** (a context outside the family home is changed) (3)	pr_or_en
Program aim (PA) *What is the aim of the program according to the author?*		Take into account all aims that the author explicitly mentions concerning the conducted program (with formulations like "the program intends to", "aims at", "wants to", "focuses on", we [the authors] examined" or "the purpose of the study..." and the like). If the author didn't explicitly mention a goal for the specific program that was conducted, BUT did formulate directed hypotheses, assess these as aims. If the author neither did explicitly mention a goal for the specific program that was conducted nor formulated directed hypotheses, BUT reports aims of a broader strategy (e.g. in a study about the effects of the parenting program "Familias Unidas" the author mentions that parenting programs in general are designed to…), check "not explicitly mentioned" and assess these aims. If problem behavior in general is reported, check the box for preventing internalising and for externalising problems, as well.	
		☐/☒ **Not explicitly mentioned**	aim_nem
	1	☐/☒ **Language promotion** (Language of the study country, the home country or bilingualism, e.g. reading or writing competencies, telling stories)	aim_c_lang
	1	☐/☒ **Cognitive promotion** (e.g. (meta)cognitive competencies, promote inductive reasoning, decision making)	aim_c_cog
	1	☐/☒ **School success improvement** (e.g. school performance, school dropout, higher education)	aim_c_scsu
	1	☐/☒ **(School) motivation enhancement** (e.g. school bonding/attendance, college enrollmt., mastery motivation, participation)	aim_c_scmo
	1	☐/☒ **Promotion of social/social-cognitive competence** (e.g. social skills, empathy, moral reasoning, social-problem solving, self-control, adaptive functioning)	aim_c_soc
	1	☐/☒ **Emotion regulation enhancement**	aim_c_emo
	1	☐/☒ **Promotion of self-esteem**	aim_c_se
	1	☐/☒ **Ethnical identity** (sense of belonging)/inter-group relation support	aim_c_ethn
	1	☐/☒ **Prevention of externalising problems** (e.g. anti-social behavior, violence, aggressiveness)	aim_c_ext
	1	☐/☒ **Prevention of internalising problems** (e.g. depression, anxiety, stress)	aim_c_int
	1	☐/☒ **Pregnancy/sexual prevention**	aim_c_sex
	1	☐/☒ **Substance abuse prevention**	aim_c_suab

	1	☐/☒ **Job training/professional success**	aim_c_job
	1	☐/☒ **General developmental support** (generally rather applies to young aged children, school readiness, motor skills)	aim_c_gd
	1	☐/☒ **Other child/adolescent related aim** (1)	aim_c_o
	2	☐/☒ **Child care improvement** (meeting physical needs of the child, like feeding competence)	aim_p_chca
	2	☐/☒ **Parenting skill enhancement** (e.g. constructive parenting behavior, parent-child communication)	aim_p_psk
	2	☐/☒ **Reduction of violence in child-rearing**	aim_p_vio
	2	☐/☒ **Parent-child relationship improvement** (e.g. warmth, bonding, parent involvement/investment, attachment)	aim_p_crel
	2	☐/☒ **Family communication/family relationship modification** (e.g. communication style of the whole family)	aim_p_fcom
	2	☐/☒ **Social support/social network enlargement of parents/family**	aim_p_sosu
	2	☐/☒ **Academic/cognitive competence promotion** (e.g. parent as teacher)	aim_p_acco
	2	☐/☒ **Parents' language competence promotion**	aim_p_lang
	2	☐/☒ **Parental professional success**	aim_p_prof
	2	☐/☒ **Promoting the process of acculturation**	aim_p_accu
	2	☐/☒ **Other parent/family related aim** (2)	aim_p_o
	3	☐/☒ **Teacher career-related skills promotion**	aim_e_tsk
	3	☐/☒ **Peer relation modification/social network within classes/school** (e.g. classroom climate)	aim_e_peer
	3	☐/☒ **Neighborhood climate/network/contact enhancement**	aim_e_neigh
	3	☐/☒ **Curricular changes/school type change/special school lessons**	aim_e_curr
	3	☐/☒ **Other environmental related aim** (3)	aim_e_o
PA: other 1	Fill out only if "Program aim" = Other child/adolescent related aim, otherwise leave blank		aim_c_otxt
PA: other 2	Fill out only if "Program aim" = Other parent/family related aim, otherwise leave blank		aim_p_otxt
PA: other 3	Fill out only if "Program aim" = Other environmental related aim, otherwise leave blank		aim_e_otxt
Addressee *Who was the program subject?*	☐/☒ **Child/Adolescent related content (1)**		addr_ch
	☐/☒ **Parent Training/Family content (2)**		addr_pf
	☐ **mother** (primarily mothers in the sample, i.e. more than 70% of the sample) ☐ **father** (primarily fathers in the sample, i.e. more than 70% of the sample ☐ **both** (between 31% and 69% mothers/fathers in the sample or both parents/caregivers are involved in the program		addr_p_spec
	☐/☒ **Environmental change (3)** (e.g. teacher promotion, institutional/system change)		addr_en
Program content (PC) *Assess the content of the program or its components, respectively.*	1	☐/☒ **Language content** (In the program linguistic and meta-linguistic skills are trained, which means improving the ability of producing, recognising, understanding, interpreting or translating written or spoken language (e.g. reading, writing, vocabulary or speech programs).)	co_c_lang
	1	☐/☒ **Learn or cognitive content** (In the program participants work on their learning style and learning success itself. Included are also programs which focus on the meta/-cognitive development or the achievement motivation, e.g. math, science, school motivation, IQ, or replacement programs.)	co_c_lecog
	1	☐/☒ **Anti-victimisation content** (In the program anti-victimisation strategies (psychological and physiological) are being trained and the consequences caused by victimisation, like traumatisation, are being broad up and worked on, e.g. child abuse program.)	co_c_vict

	1	☐/☒ **Socio-emotional competence content** (In the program life skills and socio-emotional competences, such as strategies to keep up a healthy psycho hygiene and functional social relations are trained, e.g. conduct problems, social competence, emotional well-being, self-esteem.)	co_c_soce
	1	☐/☒ **Substance abuse content** (In the program the reduction of the consumption of illegal substances is trained, e.g. alcohol or smoking program.)	co_c_suab
	1	☐/☒ **Sexual behavior content** (In the program the modification of sexual behavior is being picked out as a central theme, e.g. births control or sexual education program.)	co_c_sex
	1	☐/☒ **Career content** (In the program participants' careers are modified, e.g. work program, career coaching)	co_c_car
	1	☐/☒ **General development content** (In the program the child's general (early) development is being trained, e.g. school readiness, early education or motor skills programs.)	co_c_gd
	1	☐/☒ **Other child/adolescent related content** (1)	co_c_o
	2	☐/☒ **Parenting skill content** (In the program constructive and consistent parenting or communication behavior between parents and their children are trained, respectively.)	co_p_psk
	2	☐/☒ **Family communication/relation content** (In the program generally functional family communications and the improvement of family relations are being discussed, e.g. working on establishing constructive communication patterns between parents/partners.)	co_p_fcom
	2	☐/☒ **Parents' academic and cognitive content** (In the program parents' academic and cognitive abilities, including job related competences, are trained.)	co_p_accog
	2	☐/☒ **Parents' language competence content** (In the program parental language abilities are trained.)	co_p_lang
	2	☐/☒ **Parental social support content** (In the program parents learn to contact sources that support them, e.g. other parents, teachers, doctors.)	co_p_socsu
	2	☐/☒ **Other parent/family related content** (2)	co_p_o
	3	☐/☒ **Teacher training content** (In the program pedagogical skills of the teacher are trained, e.g. teacher knowledge, didactic abilities.)	co_e_tt
	3	☐/☒ **Peer communication content** (In the program the interaction/relation between peers/classmates and the modification of social networks within classes/schools, including classroom climate, are worked on.)	co_e_peer
	3	☐/☒ **Neighborhood enhancement content** (In the program the neighborhood's climate and the contact within/between neighborhoods are picked out as central themes, e.g. new community centre with family activities being offered, socialisation groups)	co_e_neigh
	3	☐/☒ **School structure content** (In the program the curriculum or school type of a school is changed, e.g. introducing special school lessons, day care program, offering additional schooling.)	co_e_scst
	3	☐/☒ **Other environmental related content** (3)	co_e_o
PC: other 1	Fill out only if "Program content" = Other child/adolescent related content, otherwise leave blank		co_c_otxt
PC: other 2	Fill out only if "Program content" = Other parent/family related content, otherwise leave blank		co_p_otxt
PC: other 3	Fill out only if "Program content" = Other environmental related content, otherwise leave blank		co_e_otxt

Program type *What type/kind of program has been used?* **! This variable defines the relevant program characteristics for coding the comparison level !**	☐/☒ **Nursery-program** (A child around the age of one is the primary subject.)	prt_nurs
	☐/☒ **Early education/prevention** (A child from about the age of one up to about the age of three is the primary subject.)	prt_eaed
	☐/☒ **Child training program** (A child from about three years of age up to about the age of twelve is the primary subject.)	prt_chtp
	☐/☒ **Adolescent training program** (An adolescent between thirteen and eighteen years of age is the primary subject.)	prt_adtp
	☐/☒ **Preschool/School curriculum** (A curricular program for children from three years of age.)	prt_prsc
	☐/☒ **Mentoring program** (e.g. Big brother/big sister program, tutoring)	prt_ment
	☐/☒ **Parent training** (including parent-as-teacher)	prt_pt
	☐/☒ **Family counseling**	prt_fcou
	☐/☒ **Psychotherapeutic prevention**	prt_psyt
	☐/☒ **Teacher training**	prt_tt
	☐/☒ **Schooling concept** (e.g. special school type, establish a school)	prt_scco
	☐/☒ **Transition program** (e.g. change of school, relocation)	prt_tran
	☐/☒ **Job program** (e.g. job support or training)	prt_jobp
	☐/☒ **Other**	prt_o_c
Program type: other	Fill out only if "Program type" = Other, otherwise leave blank	prt_o_tx

TECHNICAL PROGRAM INFORMATION		
Theoretical foundation *To which theoretical framework does the author primarily refer to (Why and how should the program work?)*	1 = **General developmental theory**: Child's *general* development (like motor development) is described by stages through which children typically progress (e.g. Erickson, Flavell, Griffith) 2 = **Cognitive developmental theory**: Child's *cognitive* development is described by stages through which children typically progress (e.g. Piaget) 3 = **Social learning theory**: Underlines the importance of social and cultural factors in the child's development (e.g. Bandura, Vygotsky, Forgatch) 4 = **Cognitive-behavior theory**: Based on the premise that a cognitive change results in a behavioral change (e.g. Meichenbaum) 5 = **Ecological-systemic theory**: the development of a child is influenced by factors within itself and by environmental influences (e.g. Bronfenbrenner, Brandstädter, Learner, Sameroff & Chandler, Bertalanffy) 6 = **Poverty/Migration specific theory** 7 = **Other** (also name theory if it can be assumed that a theory has been applied, however it is not well-known or labeled, e.g. theory on reading mastery)	theo_fo
Theoretical foundation poverty/migration: other	Fill out only if "Theoretical foundation" = 6, otherwise leave blank Describe the theory and name the subgroup (poverty or migration)	theo_fo_pm
Theoretical foundation: other	Fill out only if "Theoretical foundation" = 7, otherwise leave blank	theo_fo_o
Program method *Which method(s) did the author use (explicitly reported)?*	☐/☒ **Information transfer** (e.g. information leaflet, books, or instruction as intervention)	pm_info
	☐/☒ **Psycho-education** (explaining to encourage understanding, rather than just giving information)	pm_psyed
	☐/☒ **Incentives** (e.g. token system, vouchers, reward)	pm_incen
	☐/☒ **Behavioral training with instruction** (modeling, rather behaviorally oriented)	pm_beht
	☐/☒ **Video-feedback**	pm_vidfb
	☐/☒ **Group discussion** (also assistance of peers, e.g. as aid to executor, peer mediation)	pm_grpeer
	☐/☒ **Exercises** (rather cognitively oriented e.g. problem task, repetition exercise, writing essays, also play-like exercised, e.g. number games, rhyming, songs, home work/keep a diary)	pm_exhowo
	☐/☒ **Process diagnostics** (e.g. adapt the program to the participant, considering the child's or parent's abilities)	pm_pdiag
	☐/☒ **Free or guided play** (e.g. in the context of play therapy or pedagogically guided, trading games, sport, dance, movement, yoga)	pm_ phplay
	☐/☒ **Home visitation**	pm_hovis
	☐/☒ **Enriching materials** (material placed at children's or parents' disposal, e.g. books or toys, handouts with age appropriate activities, multimedia, videos, webpage)	pm_ mmenr
	☐/☒ **Performance** (e.g. through a film, theater, role play, reading or learning together, building toys together, drawing/painting)	pm_ropper

	☐/☒ **Internship/work sample** (e.g. job opportunity)	pm_insh
	☐/☒ **Joint activities** (helps create a common context of experience between student and teacher or parents apart from everyday interactions, e.g. go swimming together, field trips)	pm_jact
	☐/☒ **Networking** (e.g. mutual support of teenage mothers)	pm_netw
	☐/☒ **Other**	pm_o_c
Program method: other	Fill out only if "Program method" = Other, otherwise leave blank	pm_o_tx
Manualisation *Was a manual used for the program?*	0 = **No** 1 = **Probably** (Whether a manual was used, has not been explicitly mentioned, but is very probable, e.g. a pre-version of a manual has been mentioned or a curriculum has been used) 2 = **Yes** (The usage of a manual was explicitly mentioned.) Note: A curriculum rather consists of rough guidelines concerning the program, while a manual contains concrete information concerning the program, like contents, methods used or time planning.	manual tested: 0 vs. 1,2
Structuredness *To which degree is the program in terms of content and methodical-didactically determined?*	1 = **Very low** (All open offering programs and programs with optional attendance, e.g. programs that use self-help materials (books/films)) 2 = **Low** (Largely free composition of content and program methods/didactics, at most largely the same number/duration (intensity) or regularity of the meetings/sessions/units) 3 = **Medium** (Regularity of the contents and/or program methods/didactics, which go beyond the intensity/regularity in "low structuredness", e.g. wide predetermination of topics without predetermining the exact procedure or sequence of the topics) 4 = **High** (Detailed predetermination of contents and methods/didactics mostly realised through concrete manuals, e.g. exact procedure or sequence of the topics are predetermined)	struct tested: 1,2 vs. 3,4
Individualisation *The degree to which the programs' content and methods/didactics are adaptable to individuals' needs or skills.*	1 = **Very Low** (No recognisable individualisation. Applies for example to programs which are exclusively implemented through films.) 2 = **Low** (Everyone gets the same program, however it is likely that different skill levels/needs of the people were taken into account, e.g. if group sizes are low.) 3 = **Middle** (Some adaptation of the contents/methods, like different strategies are proposed depending on the recipient or there is a specification based on group characteristics, e.g. gender, age.) 4 = **High** (Individualising as an explicit feature of the program, e.g. after systematic diagnostics certain contents are assigned, intervention components are administered based on skills/abilities/other personal traits.)	individ tested: 1,2 vs. 3,4
Delivery format *What is the format of delivery of the program?*	1 = **Self-help** 2 = **Individual** (the program is being delivered one-by-one, e.g. individual counseling) 3 = **Group** (the program is being delivered in a group setting, e.g. small group, entire classroom) 4 = **Both** (individual and group delivery are being combined, e.g. a combination of individual mentoring and whole class teaching)	del_form

Level of participation *The degree to which the participants take actively part in the program.*	1 = **Low** (Program is primarily based on the frontal transfer of knowledge, e.g. only lecture is given or video/film shown) 2 = **Middle** (Some active participation is included, e.g. frontal transfer of knowledge combined with homework given, individually working on exercises) 3 = **High** (Participants take active part in the program by e.g. discussions, role play, behavioral training, or can influence the procedures of the intervention)		l_partic tested: 1,2 vs. 3
Culturally tailored	0 = **No** (If no program characteristic is adapted to the participant's culture; also if only the measurement of the dependent variable is available in the demanded language) 1 = **Yes** (Author explicitly states, that the program is "Culturally tailored" or adapted to the ethnic/cultural/migrant group studied or the program addresses one of the following: Ethnic identity/identity conflict, cultural peculiarities or the acculturation gap between parents and child, official language of the host country as a foreign language or mother tongue of the immigrant group (if different from the host countries official language) or ethnic matching of the service delivery agent (mentor, tutor, trainer, counselor etc.) to the immigrant group); for example the executor speaks Spanish/English or program exercises are offered in English/Spanish depending on the participant's needs		cult_tai
Implementation control	Note: If not explicitly reported, check "no". nip = not in person; program has not been conducted in person, but by video tape or written information only. The variables training, supervision, observance and additional services, strongly depend on how extensive the program itself was, i.e. for a simple and short program a one hour training/supervision/observance session might be totally sufficient (ergo: rate full). For orientation: over 10% is considered as full.		
	☐nip/☐no/☐basic/☐ full	**Training of the executor** (*How intensive was the training of the executor?*) - **Basic**: The executor received some training, like basic instructions or an information sheet - **Full**: The executor received an extensive training for conducting the program, e.g. several training sessions with feedback, or already conducted the program in the past. Note: If the executor is the developer of the program and no further information are reported, than code "full".	ic_train tested: no vs. b/f
	☐nip/☐no/☐basic/☐ full	**Supervision during program** (*How intensive was the supervision during the program? Rate the quality of supervision rather than the quantity*) - **Basic**: Some supervision has been conducted, e.g. a small amount of all program sessions were observed and evaluated/discussed with the executor or there were meetings once in a while/monthly were the executor(s) could discuss conduct problems with a supervisor - **Full**: The supervision during the program was extensive, e.g. meetings on a regular basis to discuss program methods, video recordings or discussions about program conduct with a supervisor	ic_supvi tested: no vs. b/f

	☐nip/☐no/☐basic/☐ full	**Observance while program conducted** (*To which degree has the executor been observed while conducting the program? Rate the quantity of observations*) - **Basic**: Once in a while the executor has been observed while conducting the program, e.g. at the beginning of the program random sessions have been observed - **Full**: Extensive observance of how the executor conducts the program, e.g. video recordings or personal observance of many sessions	ic_ obcon tested: no vs. b/f
	☐no/☐basic/☐ full	**Documentation of program conduct** (*How well was the program conduct documented?*) - **Basic**: Some documentation is mentioned, e.g. some sessions have been recorded on video in order to check program fidelity, in some sessions records have been kept or some sessions were observed and documented (e.g. with a checklist) - **Full**: Extensive and detailed documentation of the program conduct, e.g. many session have been recorded and analysed	ic_dcond
	☐no/☐basic/☐ full	**Documentation of study subjects' participation** (*How well was the study subject's participation documented?*) - **Basic**: The study subject's activity/on-task behavior has been rudimentarily documented, e.g. someone randomly kept record of the study subjects' compliance, eye contact or if they work quietly while doing exercises - **Full**: Extensive and detailed documentation of the study subjects' activity/on-task behavior, e.g. many sessions has been recorded and analysed regarding the study subject's compliance, eye contact or if they work quietly while doing exercises	ic_dpart
	☐no/☐basic/☐ full	**Documentation of participants' attendance** (*How well was the participants' attendance documented?*) - **Basic**: The participants' attendance has been rudimentarily documented, e.g. for each session someone kept record how many participants attended but not who - **Full**: Extensive and detailed documentation of the participants' attendance, e.g. a detailed protocol or checklist was kept about the attendance of each participant	ic_datd
	☐no/☐basic/☐ full	**Offering additional services** (*To which degree were additional services offered in order to increase or insure program compliance?*) - **Basic**: Some services were offered, like a small amount of money as expense allowance - **Full**: Many services were offered, like money, car services, child care	ic_dadse
	☐no/☐basic/☐ full	**Implementation problems** (*To which degree were implementation problems mentioned?*) - **Basic**: Implementation problems can be assumed but are not explicitly mentioned. Assumption can be based on: participants didn't pay attention to the executor, sometimes participants didn't attend the program sessions, high drop-out rate. - **Full**: Implementation problems have been explicitly mentioned, e.g. parts of the program couldn't be conducted as planned, participants' attention was hard to get or many participants didn't attend the sessions	ic_prob

INTENSITY OF THE PROGRAM		
Length of program	In days; Theoretically planned period of time over which the program will span (if reported in weeks x 7, if reported in months x 30, if reported in years x 365); school year in the USA: 300 days (36 weeks); estimation is allowed	leng_pr
Number of sessions	Number; Theoretically planned number of times the program will take place, e.g. 300 sessions; If conducted daily and reported in working weeks: x 5; …and reported in months: x 20; …and reported in years: x 200; If conducted weekly and reported in months: x 4; …and reported in years: x 45; school year in the USA: 180 instructional days; estimation is allowed	N_ses
Length of each session	In minutes; Theoretically planned duration of each program session; in case of different session lengths, type in the average length; estimation is allowed	leng_ses
Booster sessions used	0 = **Not used** 1 = **Used**	boost_ses
Frequency of sessions	1 = **Several times a day** 2 = **Daily** (on working day basis, i.e. 5 days per week) 3 = **Several times a week** 4 = **Weekly** 5 = **Several times a month** 6 = **Monthly** 7 = **Irregular or no frequency** (e.g. only one session)	freq_ses
GEOGRAPHICAL PROGRAM INFORMATION		
Area of program *In which area did the program take place?*	1 = **Rural/Suburban** (e.g. village, small town, periphery of a city) 2 = **Urban** (e.g. city center) 3 = **Mixed** (e.g. participants from city X and its catchment area) Note: If not explicitly mentioned, use recruitment area or other similar hints.	area_pr
Site of study	1 = **Single site study** (The programs takes place within a small area, e.g. one school, two communities of the same city) 2 = **Multisite study** (The programs is being conducted in a wide geographical area, e.g. in two states of the USA, in two federal stated in Germany)	site_st
Socially deprived area *Is the program being conducted in an underprivileged district?*	0 = **No** 1 = **Yes** (for example in a ghetto, social hot spot, low-income areas or a socio-economically deprived area, black or Hispanic neighborhood, socially deprived school-district) Note: Also rate "no", if no information concerning this variable is mentioned	socdepra
Setting of program *In which setting did the program take place?*	1 = **Educational facility** (e.g. preschool, kindergarten, nursery, school) 2 = **University/research facility or company** 3 = **Home** 4 = **Community centre** (e.g. church) 5 = **Hospital or special education unit** 6 = **Independently arranged setting**	set_pr

	7 = **Social Service Agency** (e.g. WIC centre, Upward Bound Centre, Head Start Centre) 8 = **Several settings** (e.g. home visitation and school setting) 9 = **Other**	
Setting of program: other	Fill out only if "Setting of program" = 9, otherwise leave blank	set_pr_o
INFORMATION ABOUT THE EXECUTOR AND DEVELOPER OF THE PROGRAM		
Predominant professional orientation of executor *Who executed the program in terms of profession?*	1 = **Not in person** (e.g. program delivery by video tape or written information only) 2 = **Educational staff** (people having educational experiences in working with children, e.g. kindergarten or school teacher) 3 = **Psychosocial professionals** (people with work experience in the psychosocial field, e.g. sociologist, psychologist, pedagogue, social worker) 4 = **Layperson or volunteers** (people with no professional experience in educating children, e.g. pensioner) 5 = **University staff** (e.g. university students as tutors) 6 = **Clinical professionals** (e.g. nurse, pediatrician) 7 = **Administrators/counselors** 8 = **Multiple personal** (several executors with different specialisations)	prof_exe
Relationship executor *Who executed the program in terms of conflict of interest?*	1 = **Not in person** (e.g. program delivery by video tape or written information only) 2 = **Third party executes the program** (e.g. kindergarten or school teacher) 3 = **With the developer loosely associated person executes the program** (e.g. developer trains third party executor, like clinicians) 4 = **Developer of the program or closely associated people** (e.g. research assistants, Ph.D. students) **take part in executing the program** 5 = **Study author(s) has/have a close relationship to the executor** (e.g. his/her research assistant, Ph.D. student, executor receives a training certificate from the author) 6 = **Study author(s) is/are the executer(s)/one of the executors** (code also 6, if the executor is the study author AND the developer)	int_exe
Relationship developer *Is there an association between the program developer and the study author?*	0 = **No conflict** (Program developer is not the study author/one of the study authors or advertised the program) 1 = **Light conflict** (The developer has a close relationship to the study author or with the author associated people (e.g. research students) 2 = **Strong conflict** (The study author took part in developing the program or advertising it. If the author(s) doesn't/don't mention the developer's name, conclude the author(s) him/herself as the developer.) Note: If there is no hint concerning the developer of the program, assume that the author is the developer.	int_dev

INFORMATION ABOUT THE CHARACTERISTICS OF THE SAMPLE

→ *FOR ASSESSING THE SAMPLE VARIABLES, USE THE TIME POINT BEFORE THE PROGRAM STARTED AS REFERENCE POINT*

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Type of group	1 = **Poverty,** 2 = **Migration** 3 = **Both**	type_gr
POVERTY AND ETHNICITY 70% of the sample needs to meet at least one or more of the 5 following poverty types. The information need to be explicitly quantified in the paper.		
Percent poverty	Percentage of the sample living in poverty	pov_percent
Income poverty	0 = **Not Applied** 1 = **Applied – parental** 2 = **Applied – paternal** 3 = **Applied – maternal** 4 = **Applied** (without hint about whose income the author assessed) The annual parental/paternal/maternal income falls below the poverty level, which can be found in appendix C - table A. If the family income is reported, divide it by 2,1 (equivalence size for an average family with two parents and two children: 1+0,5+0,3+0,3). If the threshold is not available for a certain year, use the next available estimator, e.g. if the study is published in Turkey in 1995, the best estimated threshold is 1.283US$ (average in the time span 2000-2011 in Turkey). Use alternative sources if the table does not contain sufficient information (see appendix C - table B). If the mean or median income is reported, consider the standard deviation or maximum income, respectively, in order to estimate the income level of at least 70% of the sample: - Mean (or Median) and SD are reported: $income_{70} = M + SD$ - Mean (or Median) and maximum income is reported: $income_{70} = (M + income_{max}) / 2$ If only the mean or median are reported, no estimation for the income of 70% of the sample can be made.	inc_pov
Occupation poverty	0 = **Not Applied** 1 = **Applied – parental** 2 = **Applied – paternal** 3 = **Applied – maternal** 4 = **Applied** (without hint about whose occupation the author assessed) The pursued occupation is either a low-level profession, e.g. unskilled (i.e. porters, dustmen, laborer), semi/partly-skilled (i.e. postal workers, plant/machine operators) or non-professionals or no occupation/very few hours of occupation, e.g. 10h per week is accessible, e.g. occupation category 6 und 7 of the Warner Scale (6: Assembly Line Workers -auto and other; Attendants-gas stations, hospitals; Auto Body Repairmen-painting; Baggage Men; Bank Messengers; Body and Fender Workers; Car Deliverers; Car Inspectors, railroad; Carpenters' Assistants; Chauffeurs; Chemical Plant Workers; Clerks-duplicating, fleet-service, general, receiving room. mail and shipping, stock, order, supply, postal; Coffee Roasters; Construction, iron workers; Crate Followers; Delivery Men; Drivers, motor vehicles; Electrical Workers; Exterminator's Helpers; Factory Workers; Gardeners; Gear Grinders; Golf Course Inn Keepers; Green Keepers; Guards; Jailers; Landscapers; Lathers' metallic; Lift-drivers; Machine Operators; Mail Handlers; Mailmen (letter carriers); Mechanics-gas station, building; Metal Workers-cutters, sanders; Milkmen; Moulders; Night Policemen; Night Watchmen; Operators, highway maintenance equipment; Porters, hospital; Pressers--cleaning establishment; Salesmen, route; Semi-skilled Workers;	occ_pov

	Shipping Room Workers; Shop Workers; Tank Repairman; Taxi Drivers; Draw Bridge Operators; Truck Drivers; Waiters; Waitresses; Welders; Military Servicemen; 7: Caddies; Car Washers; Construction workers; Custodians; Dairy Laborers; Dishwashers; Helpers; Heavy Laborers; House-cleaning Servicemen; Institutional Aides; Janitors; Machine Cleaners; Maintenance Man; Migrant Workers; Miners; Movers; Newsboys; Odd-job Men; Orderlies-hospital; Porters, general; Scrubwomen; Window Washers)	
Education poverty	0 = **Not Applied** 1 = **Applied – parental** 2 = **Applied – paternal** 3 = **Applied – maternal** 4 = **Applied** (without hint about whose education the author assessed) The number of years involved in the educational system falls below a certain level or a certain degree has not been achieved (see appendix D - table A. If the publication year of the study and the reference year of the threshold (see appendix D, table B) differ by more than 10 years, use alternative sources (see appendix D - table C).	educ_pov
Federal Assistance poverty	0 = **Not Applied** 1 = **Applied – parental** 2 = **Applied – paternal** 3 = **Applied – maternal** 4 = **Applied** (without hint about whose federal assistance the author assessed) Eligible for any kind of governmental aid for poverty reasons, e.g. free or reduced lunch for children, aid to families with dependent children, Medicaid, MassHealth, Hartz IV, Arbeitslosengeld, Supplemental Nutrition Assistance Program (SNAP)/Food Stamps, Temporary Assistance for Needy Families (TANF), Alaska Temporary Assistance Program (ATAP) Not automatically poverty sample: Medicare, Medi-Cal, Oregon Health Plan, Soonercare, TennCare, BadgerCare, State Children's Health Insurance Program, Program for Women, Infants, and Children (WIC), Hippy (Home Instruction for Parents and Preschool Youngsters), Head start (schools), Health Care Card (Australia)	f_as_pov
Composite poverty	0 = **Not Applied** 1 = **Applied – parental** 2 = **Applied – paternal** 3 = **Applied – maternal** 4 = **Applied** (without hint about whose federal assistance the author assessed) Any kind of system for categorising low social classes by several characteristics using sum scores, means or number of risk characteristics, e.g. High-Risk Index (based on Ramey, 2000: having at least 25/50 points equals poverty), Hollingshead Four-Factor Index of Socioeconomic Status, ad hoc measure, like meeting two out of three poverty risk characteristics (like unmarried, <12 years of education, unemployed), Schichtindex (Lampert & Kroll, 2006; 3-21, Cut-Off = 10% = 5)	comp_pov
Dominant ethnicity	1 = **White** 2 = **African (American)/Black** 3 = **Hispanic/Latino** 4 = **Diverse** 5 = **Other** Note: Dominant means more than 50% of the sample. If no group reaches 50%, code 4 (divers).	d_eth

Dominant ethnicity: other	Fill out only, if "Dominant ethnicity" = **5**, otherwise leave blank	d_eth_o
% of dominant ethnicity	Type in percentage of dominant ethnicity	p_d_eth
Status of dominant ethnicity	1 = **Minority** 2 = **Majority** (Central Intelligence Agency (US). The World Factbook. Ethnic Groups. Verfügbar unter https://www.cia.gov/library/publications/the-world-factbook/fields/2075.html#)	st_d_eth
MIGRATION Note: Sample (or subsample for which separate results are reported) have to consist mainly (at least 50%) of children/adolescents with migration background (at least one parent is born in another country). If migration background is not specified and percentage of first generation immigrants is either below 50% or not specified, but more than 90% belong to an ethnic (e.g. Hispanic or Asian American) or linguistic (e.g. English Language Learner in US) minority, that usually consists mainly of people with migration background (NOT African American, since most of them have NO migration background; also indigenous ethnic minorities are NOT included) then the inclusion criterion for migration is still fulfilled.		
Subsample of the study	1 = **Whole sample** (if whole sample fulfills inclusion criteria) 2 = **Subsample** (if only a subsample for that results are reported fulfills inclusion criteria)	subsam
Inclusion because of	1 = **1st Generation immigrants** (only if percentage is above 50%) 2 = **Migration background** (only if percentage is above 50% and percentage of 1st generation immigrants is below 50% or not specified) 3 = **Ethnic minority** (only if percentage is above 90%, migration background is not reported and percentage of 1st generation immigrants is either below 50% or not specified) 4 = **Linguistic minority** (only if percentage is above 90%, neither ethnic minority nor migration background is reported and percentage of 1st generation immigrants is either below 50% or not specified)	incl_bec
Main linguistic minorities	Name; Use the most precise information (i.e. if only English Language Learners are mentioned as participants insert "English Language Learner", but if the author also says that the home language of the participants is Spanish insert "Spanish speakers"); If more than one linguistic minority is included in the (sub-)sample insert the three linguistic minorities with the highest percentage in the sample, including the percentage of each minority in brackets and separating them by comma (e.g. "Spanish speakers (50%), Chinese speakers (40%), Portuguese speakers (10%)"); If not clear from the description of the author and "Inclusion because of" ≠ 4 then code "-99".	mig_lin
Main ethnic minorities	Name; Use the most precise information (i.e. if only Asian Americans are mentioned as participants insert "Asian Americans", but if the author also says that all the participants have Korean origin insert "Korean American"); If more than one ethnic minority is included in the (sub-)sample insert the three ethnic minorities with the highest percentage in the sample, including the percentage of each minority in brackets and separating them by comma (e.g. "Hispanic (50%), Chinese American (40%), Korean American (10%)"); If not clear from the description of the author and "Inclusion because of" ≠ 3 then code "-99".	mig_ethn
Country of origin of main immigrant groups	Name; Use the most precise information (i.e. if only Latin America is mentioned as country of origin insert "Latin America", if the author also says that all the participants are from Argentina insert "Argentina"); If more than one Country of origin is included in the (sub-)sample insert the three countries of origin with the highest percentage in the sample, including the percentage of each minority	mig_coor

	in brackets and separating them by comma (e.g. "Mexico (50%), Cuba (40%), Puerto Rico (10%)"); If not clear from the description of the author and "Inclusion because of" ≠ 1 and ≠ 2 then code "-99".	
Children/ adolescents with migration background in the sample	Percentage of children/adolescents in the sample with at least one parent born in another country; if data only is available for one parent use this percentage as best possible estimator.	mig_bgp
1. Generation immigrant children/ adolescent in the sample	Percentage of children/adolescents in the sample who were born in another country	mig_1gp
Group is predominant in the recruitment area	1 = **No, ethnic majority predominant** (in the area of recruitment the ethnic majority of the country (e.g. White in US) is predominant) 2 = **No, but minority/migrants predominant** (in the area of recruitment the minority/migrant group is not predominant but there are less majority members than minorities/ people with migration background of any origin in the area) 3 = **Yes** (the specific minority/migrant group is predominant in the recruitment area) Note: If no information about the predominant ethnicity is reported by the author but a precise (i.e. city or county, NOT only state) recruitment area within the US is given and the study was conducted after 2000 check http://www.census.gov/2010census/popmap/ipmtext.php (e.g. if you are looking for Miami City select "Florida" as State, click "areas within", choose "Places", click "Search", choose "Miami city" and click "Display". If only a precise recruitment area outside the US is given or only a precise recruitment area within the US is given but the study was conducted before 2000 ask Sebastian.	mig_pred
SAMPLE SIZE, AGE, GENDER AND RISK FACTORS (THESE VARIABLES ALWAYS CONCERN THE CHILDREN)		
N IG	Sample size at randomisation; If sample size is only reported for the whole sample, use this divided by the number of comparison groups as best possible estimator for each group.	N_IG
N CG		N_CG
Child mean age IG	In months (If reported in weeks ÷ 4, if reported in years x 12); If mean age is only reported for the whole sample, use this as best possible estimator for each group. If only the median age is reported, use this as best possible estimator for each group. If the child is not born yet, count backwards, e.g. - 4,5 months for a mothers' child in the second trimester of pregnancy For estimation based on school grade, use the following mean ages in months: Kindergarden → 66 (range: 60-72), 1st class → 78 (range: 72-84), 2nd class → 90 (range: 84-96), 3rd class → 102 (range: 96-108), and so on	M_ag_IG
Child mean age CG		M_ag_CG
Child min age IG	In months (If reported in weeks ÷ 4, if reported in years x 12); If range age is only reported for the whole sample, use this as best possible estimator for each group. If only SD is reported, subtract 2 SDs from the mean age for the minimum range and add 2 SDs to the mean age for the maximum range as best possible estimator.	min_ag_IG
Child max age IG		max_ag_IG
Child min age CG		min_ag_CG
Child max age CG		max_ag_CG
Child gender IG	In percent of females; If gender is only reported for the whole sample, use this information as best possible estimator for each group.	ge_c_IG
Child gender CG		ge_c_CG

Risk parameters of the sample: children *Is at least 50% of the sample of high risk?*	☐/☒ **Cognitive/academic developmental risk factors** (e.g. low IQ, academic problems, developmental disabilities, low birth weight)	risk_chco
	☐/☒ **Problem behavior risk factors** (e.g. conduct problems, substance abuse, criminal background/behavior, teenage pregnancy, homelessness)	risk_chpb
Risk parameters of the sample: parents *Is at least 50% of the sample of high risk?*	☐/☒ **Personal risk factors** (e.g. psychological disorders, mother's low IQ, criminal background, child abuse)	risk_pper
	☐/☒ **Structural risk factors** (e.g. teenage mother/pregnancy, divorced/single/widowed parents, inadequate prenatal care)	risk_pstr
Selectivity of the sample *To which degree did the authors select the sample?*	1 = **Low** (Sample not selective, except for inclusion criteria poverty or migration; representative sample) 2 = **Middle** (Sample selective regarding additional demographic characteristics, e.g. only boys, only children of refugees, children of divorced parents; selective program) 3 = **High** (Study participant or parents already show some problem behaviors, e.g. children of depressed mothers, children having reading problems or children rated by teachers as problematic, teenage mothers; indicated program)	sel_sam tested: 1 vs. 2,3
RECRUITMENT RATE, ATTRITION RATE AND ATTENDANCE		
Recruitment rate	In percent; Ratio of people who are willing to participate in the study/completed the randomisation process in relation to the number of people who were asked to participate/who were addressed, for example if 30 people are being called to ask for participation and 15 people take actually complete the pretest, type in "50". If the authors don't report randomisation information, use the number of participants completing the pretest. If there has been a public announcement for participation (e.g. radio announcement or advertisement on the blackboard in the school or in the newspaper) and no recruitment rate is being reported by the author, type in "-10"; If the potential participants were approached individually and no recruitment rate is being reported, type in "-30". If the potential participants were approached personally (i.e. face-to-face) and no recruitment rate is reported, type in "-40". In case of a combination of a personal/individual approach and an impersonal announcement, e.g. parent-teacher conference, type in "-20". If the sample has been methodologically preselected (e.g. interviews/screenings for preselection of a risk sample), type "-77".	recruitr
Attrition rate post: IG	In percentage; Ratio of participants (i.e. individuals who took part in the program) who didn't attend the posttest to participants who completed randomisation; If no randomisation information is available, use the number of people who completed the pretest. If attrition rate is only reported for the whole sample, use this information as best possible estimator for each group. *Note: If different sample sizes are reported for different outcomes, use the highest as sample size post.*	attr_pIG
Attrition rate post: CG		attr_pCG

Attrition rate FU1: IG	In percentage; Ratio of participants who didn't attend the follow-up I to participants who completed randomisation; If no randomisation information is available, use the number of people who completed the pretest.	attr_f1IG
Attrition rate FU1: CG	If attrition rate is only reported for the whole sample, use this information as best possible estimator for each group. If no follow-up I has been assessed, leave blank.	attr_f1CG
Attrition rate FU2: IG	In percentage; Ratio of participants who didn't attend the follow-up II to participants who completed randomisation; If no randomisation information is available, use the number of people who completed the pretest.	attr_f2IG
Attrition rate FU2: CG	If attrition rate is only reported for the whole sample, use this information as best possible estimator for each group. If no follow-up II has been assessed, leave blank.	attr_f2CG
Full attendance rate	In percent; Percentage of people attending all program sessions (-10% tolerance) If program participation is obligatory in the intervention setting, e.g. intervention is a certain school curriculum, code "-88".	atdf_r
Mean attendance rate	In percent; Mean attendance (Ratio of mean attended number of sessions to maximal number of sessions, e.g. 10 sessions planned and on average participants attended 8,7 sessions → 8,7/10 = 0,87 → attendance rate = 87%). If program participation is obligatory in the intervention setting, e.g. intervention is a certain school curriculum, code "-88".	atdm_r

VALIDITY EVALUATION

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
INTERNAL VALIDITY		
Type CG *What kind of control group was applied?*	1 = **No treatment** (The non-program group received *neither a treatment nor attention*, e.g. intensive reading program on top of the normal school curriculum versus normal school curriculum only) 2 = **Wait-list control group** (Control group will *receive the program after* the program group completed the posttests) 3 = **Treatment as usual** (Includes all control groups that are being given a *standard procedure*; e.g. counseling by a school psychologist on an as-needed basis) 4 = **Attention control group** (An *unspecific program that is unrelated* to the program's mechanism of action is conducted, in order to give the same amount of attention to the control group as the program group receives; e.g. intensive reading program versus free reading time) 5 = **Minimal treatment control group** (Control group receives a *minimal form* of treatment, e.g. one session of parent training in the control group (versus 20 sessions in the program group)) 6 = **Basic program control group** (Control group *already receives a program* component which is the same as in the program group, e.g. Head Start sample and the program group receives a parent training additionally)	type_CG
Contact between IG and CG *To which degree was an exchange between IG and CG possible concerning the content of the program strategies?*	1 = **No** (There was no contact between the IG and CG, or very unlikely) 2 = **Few** (Few contact of program strategies took place or is probable, e.g. IG and CG live in the same neighborhood, attend the same school, executor mentioned program strategies to the CG) 3 = **Some** (Some contact of program strategies took place or is very probable, e.g. IG and CG are in the same class, executor explained/practiced program strategies to/with the CG)	IG_CG tested: 1 vs. 2,3
Study design *What kind of randomisation procedure was applied?*	1 = **Cluster randomisation** (Whole clusters (e.g. class, school, site, community) are assigned to IG and CG) 2 = **Individual randomisation** (Randomisation on individual level, e.g. each child is randomised to IG and CG) 3 = **Stratified randomisation** (Participants are matched before individual randomisation or randomisation takes places within certain subgroup/strata, e.g. age, gender, grade, school, cohorts)	stud_des

Data analysis approach *How did the author handle missing values in the intervention and control group?*	1 = **Per Protocol, strong** (The authors only include participants in the data analysis if they received a certain amount of the program e.g. at least 50%) 2 = **Per Protocol, soft** (The authors include participants in the data analysis who completed the post-test, no matter how much of the program has been received; No imputation procedure was applied in order to replace missing information due to drop-outs.) 3 = **Intention to treat, soft** (The authors include all participants in the data analysis who completed the randomisation procedure, regardless of their attendance or attrition; Missing values due to drop-outs have been imputed.) 4 = **Intention to treat, strong** (The authors include all participants in the data analysis who were approached for study participation, regardless of agreement to randomisation procedure (study participation), attendance or attrition.) Note: If there is no attrition in the sample and no information about the analysis technique (per protocol or intention to treat) reported, leave blank.	dat_an_a
EXTERNAL VALIDITY		
Group equivalence *Were the groups equivalent on relevant variables prior to the program?*	1 = **Not tested** 2 = **Tested but not equivalent** (Significantly different on relevant demographic variables (e.g. age, sex, poverty/migration, primary language, risk factors) and assessment relevant variables (e.g. pretest score) 3 = **Tested and equivalent** (Not significantly different) on relevant demographic variables (e.g. age, sex, poverty/migration, primary language, risk factors) and/or assessment relevant variables (e.g. pretest score) Note: Use .05 as significance level to judge, if the groups are equivalent or not. If no tests are reported, take the author's interpretation as judgment. If subgroups are being coded and the test for group equivalence is only reported for the whole sample, use this information as best possible estimator.	gr_equiv
STATISTICAL VALIDITY		
Selectivity of outcome reporting	1 = **Not selective** (Results for all assessed outcomes that were mentioned are reported for the (sub)sample.) 2 = **Selective** (For the (sub)sample only the results for some of the assessed outcomes are reported.)	sel_outr

EFFECT SIZE LEVEL

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Publication ID		ID_Pub
Study ID		ID_Stud
Comparison ID		ID_Comp
OUTCOME INFORMATION		
Outcome	Specify outcome precisely; Use subscales if reported	outcome
Outcome category	1 = **Basic development outcomes**(BDO) 2 = **School development outcomes** (SDO) 4 = **School support outcomes** (SSO) 6 = **Parental support outcomes** (PSO)	out_cat
Outcome construct (OC)	Asses only the outcomes that can be assigned to one of the following categories, e.g. don't code physical parameters like body height or weight, heart rate or child's sicknesses. However, if a relevant outcome doesn't fit in one of the categories, discuss it in the weekly discussion group.	
	1 10 = **Cognitive development** (e.g. intelligence, reasoning, coordination, attention, distractibility, problem solving, creativity, phonological and morphological awareness, grammar, comprehension (logic), free recall) Measures: Wechsler Intelligence Scale for Children (Wechsler or WISC), Bracken Basic Concept Scale (Bracken), Stanford-Binet Intelligence Scale (SBI), Ravens Progressive Matrices (RPM), Bayley Scales of Infant Development (BSID) subscale "Mental Development Index" (MDI), Imaginative Play Predisposition Inventory (IPPI), Woodcock Johnson III Tests of Cognitive Abilities (WJ III COG), WJ Pre-School and Primary Scale of Intelligence (WPSSI) subscale "comprehension" 1 11 = **General development** (e.g. general (infant) development, motor skills, delay of gratification/impulse control, exploratory behaviour, developmental quotient, school readiness) Measures: BSID subscale "Psychomotor Development Index" (PDI), Cup Test, Denver Developmental Screening Test (DDST) subscales "Gross Motor" or "Fine Motor", Griffiths Mental Development Scale (MDS) subscale "Hand and Eye" or "Performance", Pictorial Scale of Perceived Competence and Social Acceptance (PSPCSA) subscale "Competence", Infant Behavior Scale (IBS), WJ subscale "Pre-academic skills", Bracken School Readiness Assessment (BSRA) 1 19 = **Other BDO** (e.g. combined constructs of BDO) 2 20 = **Language-related parameter** (e.g. language skills, attitudes to language, use of language, reading skills, writing skills, expressive language, spelling, analogies, listening/passage/reading comprehension, repetition, vocabulary) Measures: Peabody Picture Vocabulary Test (PPVT)-III, Test of Word Reading Efficiency (TOWRE), Test of Silent Reading Efficiency and Comprehension (TOSREC), Dynamic Indicators of Basic Early Literacy Skills (DIBLES), Woodcock Johnson-Revised (WJ-R) subscales "Passage Comprehension", "Word Identification", "Quantitative Concepts", or "Word Attack", Test of Psycholinguistic Abilities (ITPA) subscale "Total Language", Gates-MacGinitie Reading Test (GMRT)	out_con

	2	21 = **Science-related parameters** (e.g. math/science/academic achievement/skills, social science studies, arithmetic, computation, number sense, (general) knowledge, reading/writing grades, math/science grades, social science grade, GPA) Measures: Comprehensive Mathematics Inventory (CMI), WJ subscale "Mathematics Achievement" or "Applied Problems and Calculation", "Academic Skills" or "Knowledge" Number Sense Brief Total (NSB), California Comprehensive Test of Basic Skills (CTBS) subscale "Achievement",	
	2	22 = **Learning abilities** (e.g. learning strategies, learning problems, task orientation, learning motivation, academic motivation, attendance rate, days absent, learning engagement, intrinsic/extrinsic motivation, minutes spent on homework, tardiness, subjective academic abilities, academic self-concept, academic self-efficacy, educational expectations, self-empowerment) Measures: Pupil Observation Checklist (POCL) subscale "Task Orientation", Pupil Behavior Inventory (PBI) subscale "Academic Motivation", Classroom Assessment Scoring System (CLASS) subscale "Learning Formats/Engagement of Children", Attitudes Toward School Survey Questionnaire (ATSQ), Self-Regulation Questionnaires (SRQ) subscale "Learning Engagement", Academic Perception Inventory (API), Perception of Ability Scale for Students (PASS)	
	2	23 = **Career** (e.g. graduation rate, higher/further education, school biography, drop out, years of education, won an award, participation in honor society, receiving financial aid for education, special education placement, retention rate, eligibility of Early Intervention, discipline referrals, knowledge about profession, employment type, employment rate, actual employment status, job satisfaction, job prestige, owning a car, having debt, welfare dependence, attitude toward work, earning, income) Measures: Hollingshead Score	
	2	29 = **Other SDO** (e.g. combined constructs of SDO)	
	4	40 = **Teacher-related parameter** (e.g. enjoyment of teaching, personal involvement, teacher content knowledge, teacher didactic abilities/instrumental practices, teaching methods (simplification, student-centered practices), classroom organisation) Measures: Instructional Quality Assessment (IQA), Early Childhood Environment Rating Scale (ECERS-R) subscales "Personal care", "Language reasoning", "Activities", "Program")	
	4	41 = **Teacher relationship-related parameter** (e.g. teacher's sensitivity, teacher-student interaction/closeness, students' perception of teacher, support of students' learning, dealing with difficult students) Measures: ECERS-R subscales "Interactions", "Parents and staff", Preschool Classroom Rating Scale (PCI))	
	4	42 = **General (school) climate** (e.g. classroom climate, furnishings, stimulation, feeling of safety, teacher's satisfaction with school) Measures: ECERS-R subscales "Space and furnishings", CLASS subscales "positive/negative classroom climate", Early Language and Literacy Classroom Observation (ELLCO))	
	4	49 = **Other SSO** (e.g. combined constructs of SSO)	
	6	60 = **Parental attitudes & knowledge** (awareness if child's academic problems, realistic developmental/academic expectations, appropriateness of developmental expectations, parental attitude toward child's school and programs, child-centered literacy orientation, parental reading ability, reading books is one of parent's three favorite things to do	

	with their child) Measures: Parent Opinion Inventory (POI), (Mother's) Educational Attitude Scale (EAS) 6 61 = **Parental behavior** (e.g. television watching reduction, teaching behavior, applying reading techniques, e.g. vocabulary explanation, telling (bedtime) stories, visiting library, encouragement of book reading, homework check, reading out loud, having a regular reading time, number of questions asked during reading, attending school meetings, regular contact to teacher, supportiveness in play, communicative or play-like interactions with child, verbal responsivity, giving privileges for good grades) Measures: parent factor of the Parent Child Early Relational Assessment (PCERA, StimQ Cognitive Home Environment, Home Screening Questionnaire (HSQ), Parent Home Survey (PHS), Home Observation for Measurement of the Environment (HOME)) 6 69 = **Other PSO** (e.g. combined constructs of PSO)	
OC: other BDO	Fill out only if "Outcome construct" = 19, otherwise leave blank	OC_o_CDO
OC: other SDO	Fill out only if "Outcome construct" = 29, otherwise leave blank	OC_o_EO
OC: other SSO	Fill out only if "Outcome construct" = 49, otherwise leave blank	OC_o_SEO
OC: other PSO	Fill out only if "Outcome construct" = 59, otherwise leave blank	OC_o_HEO
ASSESSMENT INFORMATION		
Effect size ID		ID_EffS
Method *What kind of measurement has been applied?*	1 = **Questionnaire** (Participants read a question (not rating- or knowledge-related) and answer it in a paper-and-pencil-format, e.g. "Did you work for pay during summer?", "Have you ever had sex?") 2 = **Rating procedures** (Participant judges him/herself or another person does regarding a feature/trait by giving a rating, e.g. Child Behavior Checklist) 3 = **Achievement test** (Participant answers knowledge questions or solves problems. The questions' answers and problems' solutions are either right or wrong. e.g. IQ-test, standardised math test) 4 = **Observation** (Data is assessed through observation of relevant people, e.g. Strange situation, sensitivity of the mother playing with child) 5 = **Interview** (A (trained) person talks to the participant to gather information about the participant or an associated person, e.g. SKID, Adult Attachment Interview, questionnaire asked in interview form) 6 = **Archival data** (e.g. grades in school, number of births from hospital archives, number of incarcerations) Distinction between questionnaire and rating procedure: In contrast to questionnaires, rating procedures are test-theoretically well-founded and have a higher level of standardisation, e.g. homogenous response format.	method
Assessor *Who assessed the data (at post test)?*	1 = **Child/Adolescent** 2 = **Parent or other relative** 3 = **(Kindergarten) teacher, Director of school** 4 = **Peers** 5 = **Author/developer of the program or associated personal** (e.g. (blinded) research assistant) 6 = **Executor of the program** (If not teacher, author or developer) 7 = **Specially trained or skilled personnel** (Person is either specially trained for this individual measurement, has professional qualification for this kind of measurement (e.g. psychologist) or professional qualifications are required to administer the	assessor

	test (e.g. HAWIK – Hamburg-Wechsler-Intelligence test for children)) 8 = **Other not specially trained or skilled person** 9 = **Divers/multiple assessors** 10 = **Not in person**	
Assessment subject *Who has been assessed?*	1 = **Child/Adolescent** 2 = **Parent or other relative** 3 = **(Kindergarten) teacher** 4 = **Peers** 5 = **Multiple assessment subjects** 6 = **No person assessed** (e.g. inventory at home counted, archival data used)	ass_subj
Page	Page where inserted effect size calculation information can be found	page
EFFECT SIZE CALCULATION INFORMATION		
Assessment time point - post	Time between completion of program and post-test in month after the program is completed; Include post program time point only (program needs to be completed).	astp_p
IG Post Mean		IG_postM
IG Post SD		IG_postSD
IG Post N	If N is only reported at randomisation and exact attrition rate (AR) for post is known, calculate N under consideration of AR.	IG_postN
CG Post Mean		CG_postM
CG Post SD		CG_postSD
CG Post N	If N is only reported at randomisation and exact attrition rate (AR) for post is known, calculate N under consideration of AR.	CG_postN
IG Pre Mean		IG_preM
IG Pre SD		IG_preSD
IG Pre N		IG_preN
CG Pre Mean		CG_preM
CG Pre SD		CG_preSD
CG Pre N		CG_preN
Estimated effect size - post	Fill out only, if "Information for ES estimation - post" = 3, 4, 5, 6 or 2 in the case of change scores Note: Please insert the sample size for the estimated effect size in the fields for C1 effect size (for IG: "C1 IG Post N"; for CG: "C1 CG Post N"). If estimation is based on results of an AN(C)OVA only use F-tests with 1 numerator degrees of freedom. If an interaction factor for group (IG vs. CG) x time (pre vs. post/FU) is reported, preferably use this F-value to calculate the effect size, otherwise use the F-value for the group factor.	estES_po
Polarity - post	-1 = **Negative** (The scale to measure the outcome follows the principle: Higher values stand for impairment) 0 = **Not clear** (No direction indicated, e.g. when using opinion ratings) 1 = **Positive** (The scale to measure the outcome follows the principle: Higher values stand for improvement)	polar_ po

Information for ES estimation - post	1 = **Mean, SD and N for IG and CG <u>prior and after</u> the program are reported** 2 = **Mean, SD and N for IG and CG <u>after</u> the program are reported** (If no post SD (and post N), but pre SD (and pre N) is reported, use this as best possible estimator (Please still insert measures in the designated fields and rate missings with -99, more precisely, don't insert the preSD in postSD field even if it will be treated as such); If (in exceptional cases) comparable variables in terms of content are measured in pre-post-comparison, e.g. two versions of an IQ-test, then calculate ES from postES - preES) 3 = **Effect size is reported** (consider the index and calculation, e.g. if Cohen's d is reported, convert into Hedges' g) 4 = **Test statistic is reported** a) Mean value test (e.g. *F*, *t*) *Note: For calculations based on F-value make sure to determine direction of effect!* b) Measures of associations (e.g. *r*, χ^2) c) Proportions (e.g. *OR*) d) Non-parametric statistics (e.g. *U*) 5 = **Significance level is zero** (estimated effect size = 0; In case of a reported exact p-value, a more accurate back calculation can be made through inference on test statistics and degrees of freedom. Code 4a-d, depending on which applies) 6 = **Estimation by the coder: d = 0.2(small)/0.5(medium)/0.8(high) effect** (from information in the text, e.g. description of effects (e.g. the effects of the program where small), graphs, statistical information insufficient for back calculation etc.; please note hints on corresponding estimates) Note: Final effect size is Hedges' g Note: The coding order corresponds to the dominance of one effect size calculation information over another, e.g. if an effect size can be estimated by proportions and by a mean value test, like t-test, choose the mean value test (4a before 4d). If the authors only mentioned that an effect is significant, without giving information about the test statistic, ask for solution in the weekly discussion groups.	infES_ po
Significance - post *(explicitly mentioned)*	0 = **Effect is not significant** ($p \geq 0.05$) 1 = **Effect is significant** ($p < 0.05$) If an AN(C)OVA without repeated measures was calculated and p-values for each factor reported, use preferably significance of group x time factor, otherwise use significance of group factor as best possible estimator. If an AN(C)OVA with repeated measures was calculated, discuss in weekly discussion group.	sig_ po
Assessment time point - follow-up I	In months after program is completed	astp_fu1
IG FUI Mean		IG_fu1M
IG FUI SD		IG_fu1SD
IG FUI N	If N is only reported at randomisation and exact attrition rate (AR) for FUI is known, calculate N under consideration of AR.	IG_fu1N
CG FUI Mean		CG_fu1M
CG FUI SD		CG_fu1SD
CG FUI N	If N is only reported at randomisation and exact attrition rate (AR) for FUI is known, calculate N under consideration of AR.	CG_fu1N

Estimated effect size - FUI	Fill out only, if "Information for ES estimation - FUII" = 3, 4, 5 or 6 Note: Please insert the sample size for the estimated effect size in the fields for C1 effect size (for IG: "C1 IG FUI N"; for CG: "C1 CG FUI N"). If estimation is based on results of an AN(C)OVA only use F-tests with 1 numerator degrees of freedom. If an interaction factor for group (IG vs. CG) x time (pre vs. post/FU) is reported, preferably use this F-value to calculate the effect size, otherwise use the F-value for the group factor.	estES_fu1
Polarity - FUI	-1 = **Negative** (The scale to measure the outcome follows the principle: Higher values stand for impairment) 0 = **Not clear** (No direction indicated, e.g. when using opinion ratings) 1 = **Positive** (The scale to measure the outcome follows the principle: Higher values stand for improvement)	polar_fu1
Information for ES estimation - FUI	1 = **Mean, SD and N for IG and CG prior and after the program are reported** 2 = **Mean, SD and N for IG and CG after the program are reported** 3 = **Effect size is reported** 4 = **Test statistic is reported** a) Mean value test (e.g. *F*, *t*) b) Measures of associations (e.g. *r*, χ^2) c) Proportions (e.g. *OR*) d) Non-parametric statistics (e.g. *U*) 5 = **Significance level is zero** 6 = **Estimation by the coder: d = 0.2(small)/0.5(medium)/0.8(high) effect**	infES_fu1
Significance - FUI *(explicitly mentioned)*	0 = **Effect is not significant** 1 = **Effect is significant** Note: See significance - post.	sig_fu1
Assessment time point follow up II	In months after program is completed	astp_fu2
IG FUII Mean		IG_fu2M
IG FUII SD		IG_fu2SD
IG FUII N	If N is only reported at randomisation and exact attrition rate (AR) for FUII is known, calculate N under consideration of AR.	IG_fu2N
CG FUII Mean		CG_fu2M
CG FUII SD		CG_fu2SD
CG FUII N	If N is only reported at randomisation and exact attrition rate (AR) for FUII is known, calculate N under consideration of AR.	CG_fu2N
Estimated effect size - FUII	Fill out only, if "Information for ES estimation - post" = 3, 4, 5 or 6 Note: Please insert the sample size for the estimated effect size in the fields for C1 effect size (for IG: "C1 IG FUII N"; for CG: "C1 CG FUII N"). If estimation is based on results of an AN(C)OVA only use F-tests with 1 numerator degrees of freedom. If an interaction factor for group (IG vs. CG) x time (pre vs. post/FU) is reported, preferably use this F-value to calculate the effect size, otherwise use the F-value for the group factor.	estES_fu2

Polarity - FUII	-1 = **Negative** (The scale to measure the outcome follows the principle: Higher values stand for impairment) 0 = **Not clear** (No direction indicated, e.g. when using opinion ratings) 1 = **Positive** (The scale to measure the outcome follows the principle: Higher values stand for improvement)	polar_fu2
Information for ES estimation - FUII	1 = **Mean, SD and N for IG and CG prior and after the program are reported** 2 = **Mean, SD and N for IG and CG after the program are reported** 3 = **Effect size is reported** 4 = **Test statistic is reported** a) Mean value test (e.g. *F*, *t*) b) Measures of associations (e.g. *r*, χ^2) c) Proportions (e.g. *OR*) d) Non-parametric statistics (e.g. *U*) 5 = **Significance level is zero** 6 = **Estimation by the coder: d = 0.2(small)/0.5(medium)/0.8(high) effect**	infES_fu2
Significance - FUII *(explicitly mentioned)*	0 = **Effect is not significant** 1 = **Effect is significant** Note: See significance - post.	sig_fu2

VALIDITY EVALUATION

VARIABLE	CODING VALUES, DESCRIPTIONS AND NOTES	CODE
Standardisation *Did the author use a standardised measurement?*	1 = **Standardised measurement** (The applied measure was standardised in another (cited) study to measure the target outcomes (e.g. application of the Strange Situation (Ainsworth et al., 1978) to assess infant-mother attachment or cites Achenbach using the (corresponding scales of the) Child Behavior Checklist to measure internalizing and externalising problems) 2 = **Adapted standardised measurement** (A standardised measurement was adapted by the author for reasons of e.g. language (an English questionnaire translated for Spanish participants), cultural (different religious examples for Christian and Muslim children) or time reason (measurement is too long for the program in its original form); also code this category if a standardised interview (including e.g. a coding form and/or several trained coders) has been conducted 3 = **Ad hoc** (The measure is self-constructed. This means either no other study is cited using this measure or cited scales are modified or applied to measure another construct, e.g. using the internalising problems scale of the Child Behavior Checklist to measure externalising problems; also medical record information, such as pregnancy test; also code this category if an unstandardised interview has been conducted)	stand
Level of effect *On which level of impact is the outcome?*	1 = **Learning level** (The impact of the program on knowledge, attitudes or skills, e.g. knowledge about risks of drug-taking, drug refusal skills, attitudes towards drugs, problem solving skills, attitudes towards math) 2 = **Behavior Level** (The extent to which the acquired knowledge, attitude or skill is applied in real-life contexts and behavior outside a specific "testing situation" is changed; e.g. attending school, (frequency of) language usage, having unprotected sex within the last year, taking drugs within the last year, beating children within the last year, Becks Depression Inventory, strange situation) 3 = **Results level** (The extent to which the program affects socially desired results that are partially but not exclusively influenced by participants knowledge, attitudes, skills or behavior; e.g. grades at school, get pregnant, get a job, get incarcerated, presence of books)	loef

Proximity
How close is the measured outcome to the contents and aims of the program?
	1 = **very distal** (If none of the three below mentioned questions can be answered with "yes")
	2 = **distal** (If only one of the three below mentioned questions can be answered with "yes")
	3 = **close** (If two of the three below mentioned questions can be answered with "yes")
	4 = **very close** (If all below mentioned questions can be answered with "yes")
	Use the following three aspects to rate:
	- Is the addressee of the program the tested person?
	  - Yes, if for example a child training is conducted and the child's skill is also measured.
	  - No, if for example a parent training is conducted and the child is tested.
	- Is the outcome close to the aims of the program?
	  - Yes, if author's intention is for example to improve reading and reading ability is measured.
	  - No, if author's intention is for example the improvement of parenting skills and the aggressive behavior of the child is measured.
	- Is the outcome close to the contents of the program?
	  - Yes, if for example a social competence training is being conducted and social competence is measured.
	  - No, if for example a social competence training is being conducted and aggressive behavior is measured.
		prox

I Master Data Sheet

Master data sheet

Date: 15.02.2014

Coder: Sabrina

General information about the publication

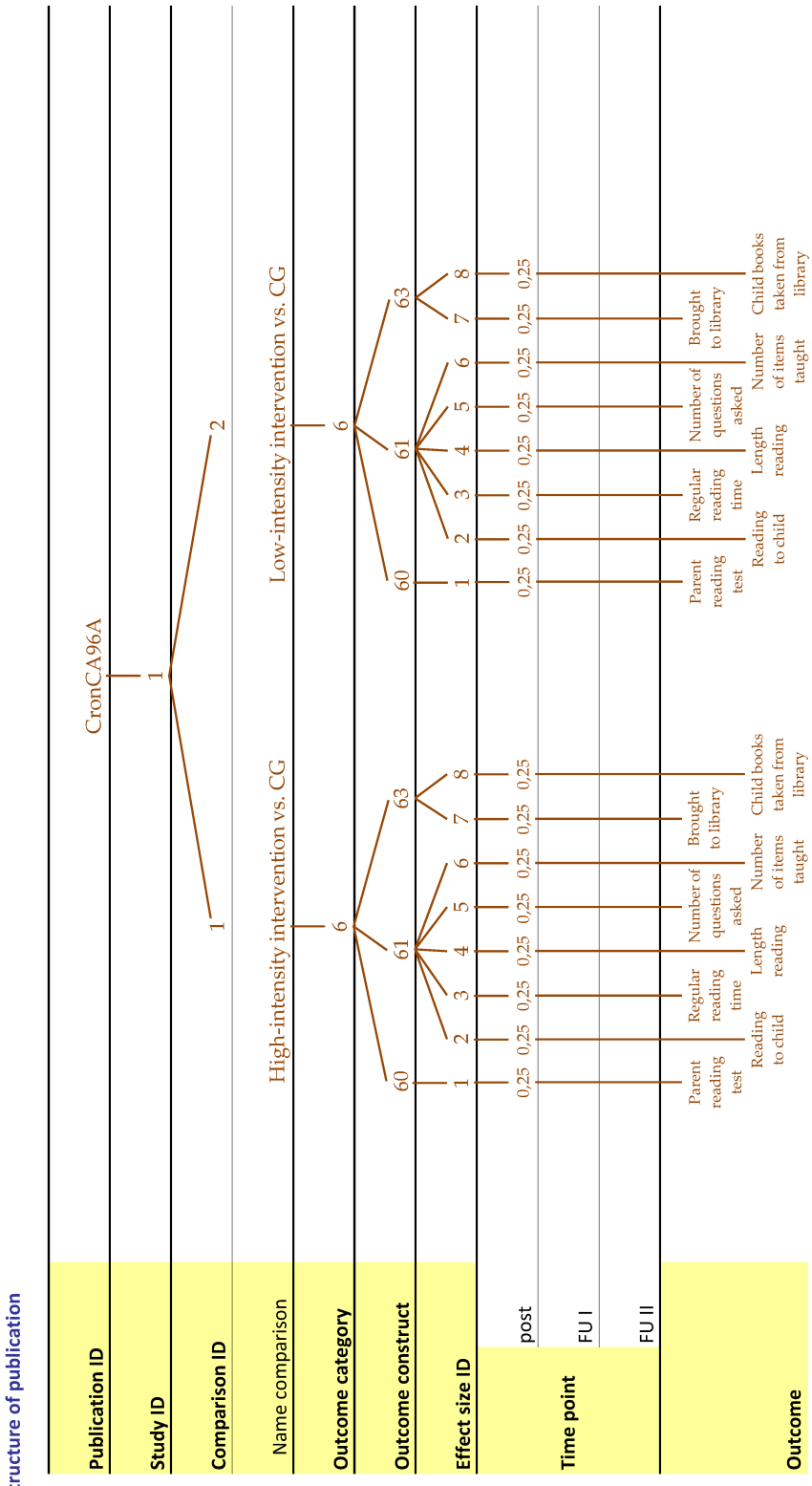
Publication ID: CronCA96A

Additional literature: 0 (0 = no, 1 = supplemental duplicate, 2 = external publication, 3 = both)

ID of additional lit.: /

Inclusion criteria

1.	Has the program been conducted with children or were children the centre of the prevention (e.g. parenting prevention)?	yes	☒	no ☐
2.	Is at least 70% of the sample socioeconomically disadvantaged?	yes	☒	no ☐
3.	Is the conducted prevention of psychosocial or pedagogic character?	yes	☒	no ☐
4.	Is the aim of the prevention the promotion of the intellectual and/or cognitive abilities?	yes	☒	no ☐
5.	Did the authors use a randomized controlled design?	yes	☒	no ☐
6.	Does each group (intervention and control group) contain at least of 25 participants? (In case of cluster-randomization, each group needs to consist of at least 5 clusters)	yes	☒	no ☐


Structure of publication
Publication ID
CronCA96A
Study ID
1
Comparison ID
1
2
Name comparison
High-intensity intervention vs. CG
Low-intensity intervention vs. CG
Outcome category
6
6
Outcome construct
60
61
63
60
61
63
Effect size ID
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
Time point
post
FU I
FU II
0,25 0,25 0,25 0,25 0,25 0,25 0,25 0,25
0,25 0,25 0,25 0,25 0,25 0,25 0,25 0,25
Outcome
Parent reading test
Reading to child
Regular reading time
Length reading
Number of questions asked
Number of items taught
Brought to library
Child books taken from library
Parent reading test
Reading to child
Regular reading time
Length reading
Number of questions asked
Number of items taught
Brought to library
Child books taken from library

J Coding Guidelines

On the following pages a selection of the coding guidelines are displayed. The instructions are roughly ordered by the hierarchical structure of the coding handbook: Basic guidelines, guidelines on publication level, guidelines on comparison level (separated in basic guidelines and variable-specific guidelines), and guidelines on effect size level.

Basic Guidelines

- **Discussion:** Discuss all unclear codings within the team.
- **Plausibility:** Always use hard facts for coding the variables. If not available use the best possible estimator.
- **Missings (not reported):** Only use *not reported* (-99) when it is unavoidable. Coding -99 is a strong statement and not an ambiguity rating. In case of insecurity, discuss variable in the discussion group.
- **Missings (blank):** Leave coding place blank, if coding does not apply.
- **Decimal places:** Concerning all numeric demographic variables: Assess two decimal places.

Guidelines on Publication Level

- **More than one study published:** If there is more than one study published in a scientific report, it will be decided in the discussion group how to handle it. If the studies seem rather independent (e.g., have different samples, are conducted in different locations), they will be coded as two independent publications. If they seem rather similar (e.g., they share a control group), they will be coded as two comparisons of one report.

Guidelines on Comparison Level: Basic Guidelines

- **More than one intervention group:** If more than one intervention group has been assessed, code comparisons for each intervention, e.g., IG1 vs. CG and IG2 vs. CG.
- **More than one control group:** If more than one control group is analysed, discuss during team meeting which one to use (most suitable). Basic rule: Least treated.
- **Whole sample versus subgroup:**
  - **Data reported:** If data is equally well reported for both, the whole sample and subgroups (e.g., regarding gender or age), code only the data for the whole sample. If considerably more variables can be filled out when coding the subgroups, choose the latter.
  - **Relevancy:** The above rule does not apply, if only (a) particular subgroup(s) is/are relevant for the research question (e.g., separate results for participants with low-SES and middle-SES).
  - **Inclusion fit:** If the whole sample of a publication suits the inclusion criteria for poverty, code this sample, even if formed subgroups suit the criteria even "better" (e.g., if the whole sample has a sample with 75% from low-SES background, it is to be preferred over a subgroup with a 90% fit).
- **Double coding:** Avoid coding the same sample more than once (Be especially attentive when coding the different Head Start studies).
- **Teenage parents:** If the sample consists of teenage parents, define and code them as parents and not as adolescents.
- **Additional programme component:** If some of the programme components were only implemented with a subgroup of the sample (e.g., a special component conducted only with a high-risk group), regard the special component as part of the programme (e.g., when coding the programme length) if the high-risk group is the majority of the sample ($\geq$ 50%). If the high-risk group is the

minority, the special component is considered as an individualization component (i.e., not taken into account when coding, for instance, the programme length).

Guidelines on Comparison Level: Variable-specific Guidelines

- **Programme type:** If possible, determine one main programme type. This defines how the rest of the comparison level following the variables *programme type* will be coded. If, for instance, a parent-as-teacher training was implemented and an additional small training component for the child itself has been conducted, consider all information when coding the variables *addressee* and *programme content*. Select *parent training* as type of programme and, from there on, neglect information regarding the child training. Only code the parent training characteristics. If an intervention with several equally substantial components was implemented (e.g., a child as well as a parent training), average the information for coding the variables on comparison level.
  - **Teacher:** Teachers are rather considered executors than addressees, because they are skilled broadcasters of knowledge and therefore act as a facilitator already.
  - **Parents:** Parents are rather considered addressees than executors, because they usually are the child's primary caregiver and hence are part of the developmental context of the child which is rather an emotional and social context than a professional one.
- **Programme method:** If parents were trained, for instance, in reading to their children and asked to practice this at home (without any observation/control by the trainer), code this as *homework*.
- **Length of programme**
  - **Booster sessions:** Booster sessions are not to be included into the length of a programme, for example, if a one year intervention was conducted followed by a booster session after six months, the length of the programme is 365 days and not 547.5.

  - **Homework:** If parents were trained, for instance, in reading to their children and asked to practice this at home (without any observation/control by the trainer), don't take the practice time into account when coding the intensity, unless the trainer is present or the parents keep record.

- **Executor:** The executor of the programme is always the one who first conducted it. This means that parents, for example, who have learned a reading strategy from a coach and intensively taught the same to their children, are not executors (but rather the subjects of the programme).

- **Child's gender:** If the gender distribution is not reported, code on the basis of plausibility, for instance, if average school classes are examined, code 50%.

Guidelines on Effect Size Level

- **Effect size**
  - **Decimal places:** Assess three decimal places for effect sizes.
  - **Effect direction:** Positive effect sizes indicate that the intervention group did better than the control group.
  - **Uncertainty calculation**: Discuss all estimated effect sizes in the discussion group.
  - **Accurate coding:** Always use the most accurate/detailed information for coding effect sizes (e.g., always use subscales instead of total test).
  - **SD and SE:** Always use standard deviations when calculating effect sizes and not standard errors. Convert if necessary.
  - **Structural equation model:** If the results are reported in a structural equation model, no effect size estimation is possible because the whole model has been tested and not exclusively the intervention-control group comparison only.
  - **Linear regression:** The regression coefficient β of linear regressions can be interpreted as estimation for the effect size if *group* (and *pretest*) is/are the only predictor(s).

  - **Clustered effect sizes:** If the effect is measured on cluster level, for instance, teacher behaviour or classroom climate, code the effects even though the sample size will most likely fall below n = 25 (the minimum number of clusters per group still needs to be n = 5, see section 2.1). Make a note in the commentary section for the corresponding effect size.

- **Assessment time point**
  - **Intermediate outcomes:** Ignore outcomes which were assessed during programme implementation. The earliest evaluation time point is directly after the intervention. Exception: If the programme continues beyond the published study, the last available measurement should be interpreted as posttest.
  - **Assessment when ending:** If the outcome assessment took place at the very end of a programme (and not exactly after), code a negative time point. This should not exceed one month. If the end of a programme is the 1st of August and the assessment takes place the 21st of July, for instance, code -0.25).
  - **Period of assessment:** If the outcome were assessed across a time span rather than a time point, average the assessment time points. For example, if the programme lasted until the end of August and the assessment was conducted from the beginning to the end of September, code 0.5.

- **Follow-up**
  - **n < 25:** Always code follow-up effects, even if the sample drops below a sample size of 25 per group. The follow-up outcome needs to fit the post outcome, otherwise code it as independent outcome with a late post assessment time point.
  - **More than three follow-ups**: If more than three follow-ups were assessed, combine the ones which were chronologically closest together (e.g., post time points: 0, 12, 24, 120, 180; combined time points: 0, 18, 150).

- **Construct Overlapping:** If possible, the outcomes were separated (e.g., by using subscales) in order to assign each outcome to the corresponding outcome construct.
- **Intellectual development and educational development:** If an intellectual outcome had to be assigned to several constructs at ones, it was assigned to the construct "General development". Such cases didn't exist for educational development.
- **Teacher and parental support:** Outcomes which could be assigned to several outcome constructs at ones, were allocated in the outcome construct "Other SEO" or "Other HEO", respectively.

K Calculation of the Between-Study Variance

$$v_i = \frac{Q - df}{C};\ df = k - 1 \tag{22}$$

$$Q = \sum_{i=1}^{k} w_i * ES_i^2 - \frac{\left(\sum_{i=1}^{k} w_i * ES_i\right)^2}{\sum_{i=1}^{k} w_i} \tag{23}$$

$$C = \sum w_i - \left(\frac{\sum w_i^2}{\sum w_i}\right) \tag{24}$$

Lebenslauf

Angaben zur Person

Name	Maichrowitz, Sabrina
Geburtsdatum	21.07.1985
Geburtsort	Frankfurt am Main
Adresse	Kernbergstraße 49, 07749 Jena
E-Mail	sabrina.maichrowitz@uni-jena.de

Ausbildung und Berfustätigkeit

1996 - 2004	Gymnasium, Marie-Curie-Gymnasium, Blankenhain Abitur 2004
2005 - 2011	Diplomstudium Psychologie, Friedrich-Schiller-Universität Jena Diplom 2011
2011 - 2016	wissenschaftliche Mitarbeiterin, Abteilung für Forschungssynthese, Intervention und Evaluation, Institut für Psychologie, Friedrich-Schiller-Universität Jena

Jena, 19.07.2016

Sabrina Maichrowitz

Ehrenwörtliche Erklärung

Ich erkläre hiermit, dass mir die geltende Promotionsordnung der Fakultät für Sozial- und Verhaltenswissenschaften der Friedrich-Schiller-Universität Jena bekannt ist.

Die vorliegende Dissertation habe ich selbst und ohne unzulässige Hilfe Dritter, insbesondere ohne die Hilfe eines Promotionsberaters, angefertigt. Bei der formalen Korrektur (z.B. Rechtschreibung, Grammatik, Stil, Syntax, Zeichensetzung) wurde ich von den folgenden Personen unentgeldlich unterstützt: Dipl.-Psych. Louisa Arnold, Dipl.-Psych. Pelle Bernhold, M.Sc. Angelika Schulz und Dipl.-Psych. Sebastian Schulz. Alle von mir benutzten Hilfsmittel und Quellen sind in der Dissertation angegeben.

Diese Dissertation wurde bisher weder als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht noch wurde eine gleiche, in wesentlichen Teilen ähnliche oder andere Abhandlung bei einer anderen Hochschule bzw. anderen Fakultät im In- oder Ausland als Dissertation eingereicht.

Die Kodierung der Studien wurde durch die Dipl.-Psych. Louisa Arnold und Dipl.-Psych. Sebastian Schulz unterstützt.

Die Dissertation wurde im Zeitraum von Oktober 2011 bis Juli 2016 unter der Supervision von Prof. Dr. Andreas Beelmann an der Friedrich Schiller Universität Jena verfasst.

Jena, 19.07.2016

Sabrina Maichrowitz